THE THEORY OF ELECTROMAGNETIC WAVES

A SYMPOSIUM

Held under the Auspices of

THE WASHINGTON SQUARE COLLEGE OF ARTS AND SCIENCE
and the INSTITUTE FOR MATHEMATICS AND MECHANICS
OF NEW YORK UNIVERSITY
and the GEOPHYSICAL RESEARCH DIRECTORATE OF THE
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES

June 6–8, 1950

1951

INTERSCIENCE PUBLISHERS, INC.

Published simultaneously in *Communications on Pure and Applied Mathematics*, A Journal Issued Quarterly by the Institute for Mathematics and Mechanics of New York University, Volume III, No. 4, pages 355–449 (1950); Volume IV, No. 1, pages 33–160, and No. 2/3, pages 225–378 (1951)

LIBRARY OF CONGRESS CATALOG CARD NUMBER 51–14016

Interscience Publishers, Inc., 250 Fifth Avenue, New York 1, N. Y.

For Great Britain and Northern Ireland:

Interscience Publishers, Ltd., 2a Southampton Row, London W. C. 1

PRINTED IN THE UNITED STATES OF AMERICA

FOREWORD

BY RICHARD COURANT

Director, Institute for Mathematics and Mechanics, New York University

With the rise of modern atomic physics, the interest of physicists in the classical theory of electromagnetism, that is, the theory based on Maxwell's equations, waned and the field was all but left to the mathematicians. At first, brilliant contributions were made by men such as H. Poincaré, Lord Rayleigh, Arnold Sommerfeld, Hermann Weyl, and G. N. Watson, who gave mathematical answers to a number of important problems of electromagnetic wave propagation. But then mathematicians too began to lose interest in a field that was becoming less and less fruitful.

At this time, when neither physicists nor mathematicians were interested in classical electromagnetism, a great number of new problems were suggested by engineers. Progress in the technique of the radio made it possible to utilize higher frequencies; further knowledge was needed of wave guide and cavity theory, diffraction and reflection of high-frequency waves, propagation in non-homogeneous atmospheres, antennas, and new principles of microwave generation.

The demands of industry coupled with military needs finally compelled the attention of scientists. The Radiation Laboratory in Cambridge was set up to attack electromagnetic problems systematically. By a new approach, rather than by the ingenious use of old tools, members of this laboratory—notably J. Schwinger and his associates—obtained far-reaching practical and theoretical results; for these results they used Green's functions, integral equations, the Wiener-Hopf procedure, variational techniques, and asymptotic methods. Mathematicians, stimulated by their success, began once again to consider electromagnetic theory a worth while field of endeavor. Even old problems, such as the influence of the ionosphere on wave propagation, a problem still incompletely solved, were attacked.

Those who attended the Symposium on the Theory of Electromagnetic Waves held at New York University in June of 1950 know how much the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories has done to support and encourage this revived interest.

In publishing the Proceedings of the Symposium, New York University and the Geophysical Research Directorate expect to increase the usefulness of the meetings. Since the beginnings of Maxwell's work in electromagnetic theory the subject has been eminently mathematical; there is thus some measure of justification in the trend to leave much of its development to mathematicians. It is hoped that the examples of research appearing in the present publication will encourage other mathematicians to cooperate with engineers and physicists in this fascinating field.

FOREWORD

BY Lt. Colonel F. C. E. ODER

*Geophysical Research Directorate, Air Force Cambridge Research Laboratories**

It is with great pleasure that the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories co-sponsored the Symposium on the Theory of Electromagnetic Waves and greets the publication of its Proceedings. As a part of its deep conviction for the necessity of long-term research, the Directorate welcomed the opportunity to encourage and accelerate research in electromagnetic theory, an understanding of which is so essential for all types of radio communication.

Indeed, the Department of the Air Force is confronted with many problems in wave propagation. In view of the tremendous range of present-day aircraft and the still greater range and altitude of future airborne vehicles, electromagnetic propagation around a spherical semiconductor surrounded by a non-concentric anisotropic inhomogeneous reflector must be considered at both large and small wavelengths.

However, we fully realize the necessity for studying fundamentals before the solutions of specific problems arising in radio communication may be obtained. Both the Geophysical Research Directorate and the Institute for Mathematics and Mechanics recognize the importance of research as distinct from development. This union of the sponsoring agencies represents on the part of each an expression of faith in the system whereby freedom of science and research must and will prevail.

A prime purpose in holding this Conference is the exchange of information within the field of electromagnetic waves; the cross-fertilization of scientific ideas in such a conference strengthens both mathematical physics and electromagnetics. In fact, the principle of the cross-fertilization of ideas is productively applied internally within the Geophysical Research Directorate. Thus, the Directorate comprises five laboratories, each dealing with a slightly different aspect of geophysics. These are the Atmospheric Analysis Laboratory, the Upper Air Laboratory, the Terrestrial Sciences Laboratory, the Atmospheric Physics Laboratory, and the Atmospheric Ionization Laboratory (formerly called the Electromagnetic Propagation Laboratory). Practices and theories used in one branch of geophysics are brought to the attention of investigators engaged in other branches of the subject. The broad field covered by the Geophysical Research Directorate includes theoretical and experimental investigations on the dynamics and structure of the atmosphere, the hydrosphere, and the lithosphere, and the relationships and energy transference among them. The various problems in each of the five laboratories mainly concern the atmosphere inasmuch as

*Now known as the Geophysics Research Division, Air Force Cambridge Research Center.

it is in this region that the Air Force operates. However, the effects occurring within the earth and the seas must also be considered because of their effect upon the lower atmosphere.

With such a broad attack on the planet as a whole, the techniques used in any one laboratory necessarily employ many scientific disciplines, including as a minimum mathematical physics, chemistry, physical chemistry, and mathematics. Only by utilizing all these various disciplines is it possible to undertake an adequate study of the mechanics and phenomena of the various divisions of the terrestrial atmosphere: the troposphere, stratosphere, chemosphere, ionosphere, mesosphere, and exosphere, the latter region extending to the boundary with interstellar space.

It is largely through the efforts of the Atmospheric Ionization Laboratory that this Symposium was initiated. Because this Laboratory employs electromagnetic waves as a probing tool, it must know the behavior of these waves in media of various types. This Laboratory uses, for example, very long radio waves to investigate the properties and characteristics of the lower ionosphere, and microwaves to examine such tropospheric mechanisms as the physics of clouds and precipitation. Similarly, the infrared and higher frequencies of the electromagnetic spectrum are utilized for a study of the ionosphere and mesosphere.

It is obvious from even this very brief indication of the activities of the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories that many problems in mathematical physics have yet to be considered. Research on these problems will undoubtedly lead to the formation of new tools for the investigation of the physics of the atmosphere.

In conclusion, I should like to express our thanks to the distinguished group of scientists who contributed to this Symposium and especially to those who crossed oceans and national borders to participate.

ACKNOWLEDGMENT

The Symposium on Electromagnetic Waves was sponsored jointly by the Geophysical Research Directorate, Air Force Cambridge Research Laboratories, and the Institute for Mathematics and Mechanics, New York University. Since recent advances in electromagnetics have been both rapid and important, and since no conference on this subject has been held for some years, the meeting was set for June, 1950. The conference was intended to bring theoreticians abreast of late developments in electromagnetics; problems still outstanding could be reviewed; and applied problems of particular urgency could be discussed. The sponsors believed that the stimulus provided by personal contacts, open discussion, and critical examination of current problems would markedly advance research already in progress.

The Symposium was conceived in the Atmospheric Ionization Laboratory of the Geophysical Research Directorate. Arrangements were made by a committee consisting of M. Kline of New York University and N. C. Gerson of the Geophysical Research Directorate, assisted by Professors S. G. Roth, B. Friedman, F. W. John, H. E. Wahlert, and Messrs. L. Kraus, J. Schmoys, and J. Lurye of New York University. Without their careful planning the Symposium could scarcely have come into being. That the Symposium was a success may be attributed solely to the cooperation, participation, and enthusiasm of the conferees.

Frederic C. E. Oder
Lt. Colonel, USAF
Director, Base Directorate for Geophysical Research
Air Force Cambridge Research Laboratories

CONTENTS

On the Theory of Electromagnetic Wave Diffraction by an Aperture in an Infinite Plane Conducting Screen

By HAROLD LEVINE and JULIAN SCHWINGER
Lyman Laboratory of Physics, Harvard University

1. *Introduction*

The diffraction of electromagnetic and light waves by an aperture in a plane conducting screen is a classical boundary value problem. As is well known, theoretical analysis aims at a solution of the vector Maxwell equations, which incorporates a prescribed form of excitation and satisfies appropriate boundary conditions on the screen and in the aperture.

A small measure of progress towards this objective results from the Kirchhoff diffraction theory, which identifies aperture and incident fields and arbitrarily assigns null values to the field components on the shadow face of the screen. The Kirchhoff formulation suitable for an electromagnetic field (assuming harmonic time variation) is given by Stratton and Chu [1]; this includes charge distributions on the rim of the screen to ensure that the free space fields obey the Maxwell equations. A defect in the Kirchhoff procedure is revealed by its failure to duplicate the assumed boundary values at the conducting screen. The lack of self-consistency has a further consequence that Kirchhoff predictions are qualitatively correct only if the wave length of the electromagnetic field is small in comparison with all aperture dimensions, for then the field on the shadow face of the screen is relatively small.

Another method of analysis, which provides information at long wave lengths, is due to Lord Rayleigh [2]. The basic idea is that, in the vicinity of the aperture, the electromagnetic field distributions can be calculated as though the wave length were infinite, making available the results of potential theory. As an example, Rayleigh treats the case of a circular aperture, with normally incident harmonic plane waves [3]. After identifying the local field with that of a Hertzian oscillator, the known radiation characteristics of the latter are used to find the diffracted field at large distances from the aperture. The tangential

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Sciences and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directories of the Air Force Cambridge Research Laboratories.

electric field at the screen vanishes in this solution, as required by the boundary condition on a perfectly conducting surface. However, the predicted transmission cross section (which measures the ratio of energy passing through the aperture per second to that transported per unit area of the incident wave) is accurate for long wave lengths only, representing the first term of an expansion for this quantity in ascending powers of the ratio, (radius of aperture/wave length). Bethe [4] considers Rayleigh's example again, and extends the theory to apply for an arbitrary spatial incident field. A new feature is the diffracted field representation by fictitious magnetic charges and currents in the aperture; their low frequency distributions are obtained, having due regard for all boundary conditions. The resulting plane wave transmission cross sections are accurate to the same order of approximation as in Rayleigh's theory.

A procedure for obtaining an exact solution to this problem, valid at all wave lengths, is described recently by Meixner [5]. The analysis is carried out with a pair of scalar electromagnetic potentials, akin to those which Debye [6] employed in the theory of diffraction by a spherical obstacle. In addition to requirements imposed by the wave equation, boundary and radiation conditions, Meixner prescribes supplementary conditions for the potentials. It is the purpose of the latter conditions, enforced at the rim of the screen, to assure quadratic integrability there for the electromagnetic field components. Indeed, the guarantee of a finite electromagnetic field energy in any arbitrarily small region of space is regarded as an essential feature of a unique and physically acceptable solution [7]. For explicit construction of the potentials, spheroidal coordinates are appropriate, as these permit a convenient description of the circular aperture, and allow separation of variables in the wave equation. Meixner obtains infinite series expansions for the potentials in terms of spheroidal functions, although numerical evaluation is deferred. In this connection, it may be anticipated that slow convergence of the series with increasing frequency will render computation difficult.

From the brief survey of theoretical methods available for three dimensional electromagnetic diffraction problems, the need for approximation procedures, accurate in a large frequency range, is evident. The latter would be particularly appropriate for problems which do not admit of analysis in terms of known solutions to the wave equation. This paper, a sequel to previous ones concerned with diffraction in a scalar field [8], describes the nature of variational principles for obtaining some of the desired information [9].

The general features of the investigation which follows pertain to the steady state diffraction problem for an aperture of arbitrary shape in a perfectly conducting screen, with incident plane electromagnetic waves.

A formal description of the fields on opposite sides of the screen, and a schedule of boundary conditions in the plane of the latter is given, utilizing symmetry properties of the Maxwell equations with respect to reflection in a plane. To apply these boundary conditions, expressions for the field vectors

within any region are derived in terms of the tangential components of either electric or magnetic fields on the boundary of the region. The mathematical tools for exhibiting such relations are tensor (or dyadic) Green's functions, whose properties are briefly described. On the far (shadow) side of the screen, the electric and magnetic field at any point can be represented by surface integrals involving the tangential electric aperture field; similar expressions apply on the near side of the screen, along with the incident and reflected fields appropriate to a completely infinite screen. The electromagnetic field thus constructed satisfies Maxwell's equations at all points of space, and moreover its tangential electric component vanishes at the screen and is continuous through the aperture. From equality of the respective tangential magnetic fields in the aperture, or equivalently, of the transmitted and incident components, an integral equation for the tangential electric aperture field is obtained. Employing the integral equation (whose solution is seldom feasible), a stationary property of the spherical wave radiation field at large distances from the aperture is established, subject to small independent variations (about the correct values) of the tangential electric aperture fields due to a pair of incident waves.

Alternatively, electric and magnetic fields on the far side of the screen are uniquely determined by values of the tangential magnetic field at the shadow face of the screen and in the aperture (the latter being equal to those of the incident magnetic field). An integral equation to determine the magnetic field distribution on the screen is a consequence of the null value there for the tangential electric field. The integral equation can be utilized to obtain another stationary property of the radiation field, which involves the distributions arising from a pair of incident waves.

A closely related variational principle is based on description of the field in terms of the current on (or the discontinuity in tangential magnetic field at) the screen. The electric and magnetic field at any point of space can be individually represented by a surface integral containing the current, to which the corresponding incident field is added. These fields satisfy the Maxwell equations and exhibit continuous variation in passing through the aperture; the requirement of vanishing tangential electric field at the screen yields an integral equation to specify the current distribution, from which the variational principle is constructed.

The plane wave transmission cross section of the aperture shares these stationary properties, based on a theorem which relates the cross section to the imaginary part of the radiation field amplitude in the direction of incidence. Complementary aspects are exhibited by the different forms of cross section, with low frequency behavior readily accessible to aperture electric field approximations, and high frequency behavior to current approximations. In general, the overall agreement of numerical results obtained from the variational formulations allows an estimate of proximity to the correct solution.

The variational method is applied in detail to the problem of diffraction by

a circular aperture, with normally incident plane waves. Numerical results for the transmission cross section are compared with those yielded by the Kirchhoff and Rayleigh approximations.

2. *Formulation of Boundary Value Problem*

We consider an infinitesimally thin, perfectly conducting plane screen S_2, of infinite extent, which is perforated by an aperture S_1, and located in otherwise empty space. A rectangular coordinate system is chosen with origin at some point of the aperture, and oriented so that the screen lies in the x,y-plane (Figure 1).

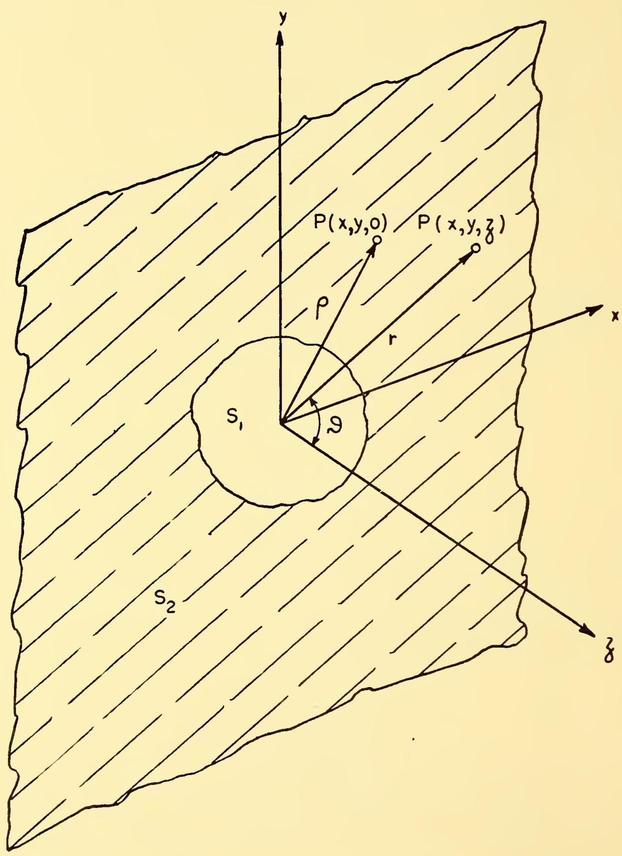


Fig. 1. Diffracting aperture in a plane screen.

A plane electromagnetic wave is incident on the aperture in the half space $z < 0$; it is desired to investigate the diffracted field. The incident wave, with propagation vector $\mathbf{n}'$ and polarization vectors $\mathbf{e}'$, $\mathbf{h}'$, is described by

$$\begin{aligned} \mathbf{E}^{\mathrm{inc}}(\mathbf{r}) &= \mathbf{e}' \exp \{ik\mathbf{n}'\cdot\mathbf{r}\} \\ &= \mathbf{e}' \exp \{ik(x \sin \vartheta' \cos \varphi' + y \sin \vartheta' \sin \varphi' + z \cos \vartheta')\} \\ \mathbf{H}^{\mathrm{inc}}(\mathbf{r}) &= \mathbf{h}' \exp \{ik\mathbf{n}'\cdot\mathbf{r}\}, \end{aligned} \tag{2.1}$$

$$\mathbf{e}' = \mathbf{h}' \times \mathbf{n}', \qquad \mathbf{h}' = \mathbf{n}' \times \mathbf{e}', \qquad \mathbf{e}'\cdot\mathbf{e}' = \mathbf{h}'\cdot\mathbf{h}' = \mathbf{n}'\cdot\mathbf{n}' = 1,$$

where $k = 2\pi/\lambda$ is the wave number and λ the wave length. The harmonic time dependence, $\exp \{-ikct\}$, with c the velocity of wave propagation ($= 3.10^{10}$cm/sec), is omitted throughout.

For the complete (incident + diffracted) field, the electric and magnetic intensities are governed by the free space Maxwell equations (Gaussian units are employed)

$$\begin{aligned} \nabla \times \mathbf{E} &= ik\mathbf{H}, \qquad \nabla\cdot\mathbf{E} = 0 \\ \nabla \times \mathbf{H} &= -ik\mathbf{E}, \qquad \nabla\cdot\mathbf{H} = 0 \end{aligned} \tag{2.2}$$

and subject to the boundary condition

$$\mathbf{e}_z \times \mathbf{E} = 0, \qquad \mathbf{r} \text{ on } S_2 \tag{2.3}$$

where $\mathbf{e}_z$ is a unit vector in the z direction; both electric and magnetic fields vary continuously through space, including traversal of the aperture.

The diffraction problem may be formulated in different ways, according to the nature of the field existence theorem employed. In addition, there are alternate geometrical viewpoints, with counterparts in mathematical formulation. For one, the aperture is obtained by excising part of a completely infinite screen, and becomes a coupling surface for the half spaces on opposite sides of the screen; the other regards the screen as an obstacle inserted in free space. Although equivalent results are obtained by the diverse procedures if the problem is treated rigorously, these lead to independent, complementary variational principles useful in the approximation sense.

Considerations of symmetry provide information about the fields on opposite sides of the screen and their relation in the aperture. Let us write

$$\begin{aligned} \mathbf{E}(\mathbf{r}) &= \mathbf{E}_0(\mathbf{r}) + \mathbf{E}_1(\mathbf{r}), \qquad \mathbf{H}(\mathbf{r}) = \mathbf{H}_0(\mathbf{r}) + \mathbf{H}_1(\mathbf{r}), \qquad z \leq 0 \\ \mathbf{E}(\mathbf{r}) &= \mathbf{E}_2(\mathbf{r}), \qquad \mathbf{H}(\mathbf{r}) = \mathbf{H}_2(\mathbf{r}), \qquad z \geq 0 \end{aligned} \tag{2.4}$$

where $\mathbf{E}_0(\mathbf{r})$, $\mathbf{H}_0(\mathbf{r})$ describe the field in the absence of an aperture,

$$\begin{aligned} \begin{pmatrix} \mathbf{E}_0(\mathbf{r}) \\ \mathbf{H}_0(\mathbf{r}) \end{pmatrix} &= \begin{pmatrix} \mathbf{e}' \\ \mathbf{h}' \end{pmatrix} \exp \{ik\mathbf{n}'\cdot\mathbf{r}\} \mp \begin{pmatrix} \mathbf{e}' \\ \mathbf{h}' \end{pmatrix}\cdot(\mathbf{\varepsilon} - 2\mathbf{e}_z\mathbf{e}_z) \exp \{ik\mathbf{n}'\cdot(\mathbf{r} - 2\mathbf{e}_z\cdot\mathbf{r}\mathbf{e}_z)\} \\ & \qquad z \leq 0 \\ \mathbf{e}_z \times \mathbf{E}_0 &= 0, \qquad \mathbf{e}_z\cdot\mathbf{H}_0 = 0, \qquad z = 0. \end{aligned} \tag{2.5}$$

Before subjecting (4) to the boundary conditions at the plane $z = 0$, it is convenient to classify solutions of the Maxwell equations according to their symmetry in the z coordinate. The even and odd solutions with respect to reflection in the plane $z = 0$ are

$$\begin{aligned} E_t(x, y, z) &= \pm E_t(x, y, -z), \qquad H_t(x, y, z) = \mp H_t(x, y, -z) \\ E_z(x, y, z) &= \mp E_z(x, y, -z), \qquad H_z(x, y, z) = \pm H_z(x, y, -z) \end{aligned} \tag{2.6}$$

respectively, where E_t, H_t signify components transverse to the z direction, i.e. tangential to the x,y-plane.

Each odd solution, whose tangential electric field components vanish in the aperture as well as upon the screen, describes a field configuration with the plane $z = 0$ completely occupied by a perfect conductor. The fields $\mathbf{E}_0$, $\mathbf{H}_0$ constitute an odd solution, resulting from superposition of the incident plane wave and a plane wave specularly reflected from the conducting surface. Owing to the geometrical identity of the half spaces $z \gtrless 0$, the fields attributed to the presence of an aperture, $\mathbf{E}_{1,2}$, $\mathbf{H}_{1,2}$, belong to the class of even solutions, viz:

$$\begin{aligned} E_{2t}(x, y, z) &= E_{1t}(x, y, -z), \qquad H_{2t}(x, y, z) = -H_{1t}(x, y, -z) \\ E_{2z}(x, y, z) &= -E_{1z}(x, y, -z), \qquad H_{2z}(x, y, z) = H_{1z}(x, y, -z). \end{aligned} \tag{2.7}$$

Accordingly, the boundary conditions relating to the tangential components of (4),

$$\begin{aligned} E_{2t} = E_{1t}, \qquad H_{2t} - H_{1t} = H_{0t}, \qquad &\mathbf{r} \text{ in } S_1 \\ E_{2t} = 0 = E_{1t}, \qquad &\mathbf{r} \text{ on } S_2 \end{aligned} \tag{2.8}$$

are satisfied if E_{2t} vanishes on the screen, and

$$H_{2t} = \tfrac{1}{2}H_{0t} = H_t^{\text{inc}}, \qquad \mathbf{r} \text{ in } S_1. \tag{2.9}$$

The boundary conditions for normal components need not be considered explicitly, as these are automatically satisfied if the conditions for the tangential components are fulfilled. We note, in particular, that

$$E_{2z} = \tfrac{1}{2}E_{0z}, \qquad \mathbf{r} \text{ in } S_1. \tag{2.10}$$

If the screen were regarded as an obstacle to the propagation of the incident wave through free space, we could write

$$\begin{aligned} \mathbf{E}(\mathbf{r}) &= \mathbf{E}^{\text{inc}}(\mathbf{r}) + \overline{\mathbf{E}}(\mathbf{r}) \\ \mathbf{H}(\mathbf{r}) &= \mathbf{H}^{\text{inc}}(\mathbf{r}) + \overline{\mathbf{H}}(\mathbf{r}) \end{aligned} \tag{2.11}$$

everywhere, subject to the boundary condition (3).

3. *Tensor Green's Functions*

Our next task is the formulation of explicit boundary value problems in accordance with the general field requirements outlined above. For this purpose, we make use of the well known existence theorem that the fields within a region are uniquely determined by the values of the tangential components of the electric field, or the magnetic field, on the bounding surface of the region. In order to exhibit this relation explicitly, the concept of the Green's function is introduced. The type of Green's function required is somewhat more general than that usually encountered, for we desire to obtain a linear relation between the field vectors within a region and the field vectors on the surface of that region. Accordingly, the coefficients in that relation must be of the character of tensors, or dyadics. The remainder of this section is devoted to the theory of the tensor Green's functions, and the derivation of field representations; our account follows closely the M.I.T. Radiation Laboratory Report 43-34(1943) by J. Schwinger.

The electromagnetic fields within a region occupied by both electric and magnetic charge and current are described by

$$\begin{aligned} &\nabla \times \mathbf{E} = ik\mathbf{H} - \frac{4\pi}{c}\mathbf{J}^*, \qquad &&\nabla \times \mathbf{H} = -ik\mathbf{E} + \frac{4\pi}{c}\mathbf{J} \\ &\nabla \cdot \mathbf{E} = 4\pi\rho, &&\nabla \cdot \mathbf{H} = 4\pi\rho^* \end{aligned} \tag{3.1}$$

where $\rho(\mathbf{r})$ and $\mathbf{J}(\mathbf{r})$ are the electric charge and current densities, and $\rho^*(\mathbf{r})$, $\mathbf{J}^*(\mathbf{r})$ represent the analogous magnetic quantities; the time dependence of all quantities is $\exp\{-ikct\}$. The electric and magnetic fields, individually, obey the equations:

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{E}) - k^2\mathbf{E} &= \frac{4\pi ik}{c}\mathbf{J} - \frac{4\pi}{c}\nabla \times \mathbf{J}^* \\ \nabla \times (\nabla \times \mathbf{H}) - k^2\mathbf{H} &= \frac{4\pi ik}{c}\mathbf{J}^* + \frac{4\pi}{c}\nabla \times \mathbf{J}. \end{aligned} \tag{3.2}$$

We shall define the tensor Green's functions associated with a region V bounded by a surface S, in terms of the field which a point current would produce within the region if it were enclosed by perfectly conducting metallic walls, coinciding with the surface S. Consider, therefore, the electric field produced by an electric current density

$$\mathbf{J}(\mathbf{r}) = \mathbf{e}\,\delta(\mathbf{r} - \mathbf{r}') \tag{3.3}$$

where $\mathbf{e}$ is an arbitrary constant vector, and $\delta(\mathbf{r} - \mathbf{r}')$ is defined by

$$\int \delta(\mathbf{r} - \mathbf{r}')\, d\tau = 1, \qquad \delta(\mathbf{r} - \mathbf{r}') = 0, \qquad |\mathbf{r} - \mathbf{r}'| \neq 0 \tag{3.4}$$

in which the integration is to be extended over a region enclosing the point $\mathbf{r}'$. The components of the electric field will evidently be linearly related to the components of the constant vector $\mathbf{e}$, and we therefore write, in dyadic notation,

$$\mathbf{E}(\mathbf{r}) = \frac{4\pi ik}{c}\boldsymbol{\Gamma}^{(1)}(\mathbf{r}, \mathbf{r}')\cdot\mathbf{e} = \frac{4\pi ik}{c}\int \boldsymbol{\Gamma}^{(1)}(\mathbf{r}, \mathbf{r}'')\cdot\mathbf{J}(\mathbf{r}'')\, d\tau'' \tag{3.5}$$

upon which is imposed the boundary condition that the tangential components of $\mathbf{E}$ vanish at the surface S, or

$$\mathbf{n} \times \mathbf{E}(\mathbf{r}) = 0, \qquad \mathbf{r} \text{ on } S \tag{3.6}$$

where $\mathbf{n}$ is the outwardly drawn normal to the surface S at the position $\mathbf{r}$. We have thereby introduced the dyadic $\boldsymbol{\Gamma}^{(1)}(\mathbf{r}, \mathbf{r}')$, which we shall call the electric field Green's function, defined by

$$\begin{gathered} \nabla \times (\nabla \times \boldsymbol{\Gamma}^{(1)}(\mathbf{r}, \mathbf{r}')) - k^2\boldsymbol{\Gamma}^{(1)}(\mathbf{r}, \mathbf{r}') = \boldsymbol{\varepsilon}\delta(\mathbf{r} - \mathbf{r}') \\ \mathbf{n} \times \boldsymbol{\Gamma}^{(1)}(\mathbf{r}, \mathbf{r}') = 0, \qquad \mathbf{r} \text{ on } S \end{gathered} \tag{3.7}$$

where $\boldsymbol{\varepsilon}$ represents the unit dyadic. Similarly, the magnetic field produced by the magnetic current density

$$\mathbf{J}^*(\mathbf{r}) = \mathbf{e}\delta(\mathbf{r} - \mathbf{r}') \tag{3.8}$$

can be written

$$\mathbf{H}(\mathbf{r}) = \frac{4\pi ik}{c}\boldsymbol{\Gamma}^{(2)}(\mathbf{r}, \mathbf{r}')\cdot\mathbf{e} = \frac{4\pi ik}{c}\int \boldsymbol{\Gamma}^{(2)}(\mathbf{r}, \mathbf{r}'')\cdot\mathbf{J}^*(\mathbf{r}'')\, d\tau''. \tag{3.9}$$

However, here the boundary conditions do not relate directly to the magnetic field, but rather to the accompanying electric field, and thus state:

$$\mathbf{n} \times (\nabla \times \mathbf{H}(\mathbf{r})) = 0, \qquad \mathbf{r} \text{ on } S. \tag{3.10}$$

The equations

$$\begin{gathered} \nabla \times (\nabla \times \boldsymbol{\Gamma}^{(2)}(\mathbf{r}, \mathbf{r}')) - k^2\boldsymbol{\Gamma}^{(2)}(\mathbf{r}, \mathbf{r}') = \boldsymbol{\varepsilon}\delta(\mathbf{r} - \mathbf{r}') \\ \mathbf{n} \times (\nabla \times \boldsymbol{\Gamma}^{(2)}(\mathbf{r}, \mathbf{r}')) = 0, \qquad \mathbf{r} \text{ on } S \end{gathered} \tag{3.11}$$

therefore define a second dyadic, $\boldsymbol{\Gamma}^{(2)}(\mathbf{r}, \mathbf{r}')$, which we shall term the magnetic field Green's function.

For infinite empty space, devoid of metallic objects, there is no distinction between electric and magnetic field Green's functions. The fundamental tensor Green's function of free space, $\boldsymbol{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}')$, which is a solution of the differential

equation (7) or (11), and describes a spherical wave moving outwards at large distances from the source point, appears in the closed form (see Appendix 1):

$$\boldsymbol{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}') = \left(\boldsymbol{\varepsilon} - \frac{1}{k^2} \nabla\nabla'\right) \frac{\exp\{ik\,|\mathbf{r} - \mathbf{r}'|\}}{4\pi\,|\mathbf{r} - \mathbf{r}'|} = \boldsymbol{\Gamma}^{(0)}(\mathbf{r}', \mathbf{r}). \tag{3.12}$$

The tensor Green's functions for a half space, with an infinite plane conducting boundary, are easily constructed. For either of the current densities (3) or (8), the fields are those in the absence of a conducting boundary, provided a suitably disposed image current is introduced. From this scheme, we find

$$\boldsymbol{\Gamma}_+^{(1),(2)}(\mathbf{r}, \mathbf{r}') = \boldsymbol{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}') \mp \boldsymbol{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}' - 2\mathbf{e}_z\mathbf{e}_z\cdot\mathbf{r}')\cdot(\boldsymbol{\varepsilon} - 2\mathbf{e}_z\mathbf{e}_z), \quad z, z' \geq 0 \tag{3.13}$$

the upper and lower signs to be employed for $\boldsymbol{\Gamma}_+^{(1)}$, $\boldsymbol{\Gamma}_+^{(2)}$ respectively. By way of verification, observe that the expression for $\boldsymbol{\Gamma}_+^{(1)}$ provides vanishing tangential electric field components at the conducting boundary ($z = 0$), whereas the normal component is double its free space value; all in accord with the boundary conditions.

In the following development, use is made of a general vector relation between surface and volume integrals,

$$\begin{aligned} &\int_S dS\,\mathbf{n}\cdot[\mathbf{B} \times (\nabla \times \mathbf{A}) - \mathbf{A} \times (\nabla \times \mathbf{B})] \\ &\quad = \int_V d\tau\,[\mathbf{A}\cdot\nabla \times (\nabla \times \mathbf{B}) - \mathbf{B}\cdot\nabla \times (\nabla \times \mathbf{A})] \end{aligned} \tag{3.14}$$

which is termed Green's second vector identity. As a first indication of its usefulness, we can show that the $\boldsymbol{\Gamma}$'s share the fundamental symmetry property of all Green's functions:

$$\boldsymbol{\Gamma}(\mathbf{r}', \mathbf{r}'') = [\boldsymbol{\Gamma}(\mathbf{r}'', \mathbf{r}')]^T \tag{3.15}$$

$$\nabla' \times \boldsymbol{\Gamma}^{(1)}(\mathbf{r}', \mathbf{r}'') = [\nabla'' \times \boldsymbol{\Gamma}^{(2)}(\mathbf{r}'', \mathbf{r}')]^T \tag{3.16}$$

where $\boldsymbol{\Gamma}^T$ denotes the transposed dyadic $\Gamma_{kl}^T = \Gamma_{lk}$. Equation (15) is established by applying Green's second vector identity to the functions $\mathbf{A}(\mathbf{r}) = \boldsymbol{\Gamma}(\mathbf{r}, \mathbf{r}')\cdot\mathbf{e}'$, $\mathbf{B}(\mathbf{r}) = \boldsymbol{\Gamma}(\mathbf{r}, \mathbf{r}'')\cdot\mathbf{e}''$, in which $\mathbf{e}'$ and $\mathbf{e}''$ are arbitrary constant vectors and the $\boldsymbol{\Gamma}$'s can be either electric or magnetic field Green's functions. The relation (16) is readily obtained from (14) by substituting

$$\mathbf{A}(\mathbf{r}) = \nabla \times \boldsymbol{\Gamma}^{(2)}(\mathbf{r}, \mathbf{r}')\cdot\mathbf{e}', \qquad \mathbf{B}(\mathbf{r}) = \boldsymbol{\Gamma}^{(1)}(\mathbf{r}, \mathbf{r}'')\cdot\mathbf{e}''.$$

The fields of physical interest are those contained within regions devoid of charge and current (equations 2.2) and both electric and magnetic fields are thus required to satisfy the vector wave equation:

$$\begin{aligned} &\nabla \times (\nabla \times \mathbf{E}) - k^2\mathbf{E} = 0 \\ &\nabla \times (\nabla \times \mathbf{H}) - k^2\mathbf{H} = 0. \end{aligned} \tag{3.17}$$

Consider now a region V' bounded by a surface S', which is contained within V, the region of definition of the Green's functions, and which may, in particular, coincide with it. We wish to express the fields within V' in terms of the tangential field components on the boundary surface S'. To this end, let us employ Green's second vector identity, with

$$\mathbf{A}(\mathbf{r}') = \mathbf{E}(\mathbf{r}'), \qquad \mathbf{B}(\mathbf{r}') = \mathbf{\Gamma}^{(1)}(\mathbf{r}', \mathbf{r})\cdot\mathbf{e}.$$

We obtain

$$-\int_{S'} dS'\, \mathbf{n}'\cdot[ik\mathbf{H}(\mathbf{r}') \times (\mathbf{\Gamma}^{(1)}(\mathbf{r}', \mathbf{r})\cdot\mathbf{e}) + \mathbf{E}(\mathbf{r}') \times (\nabla' \times \mathbf{\Gamma}^{(1)}(\mathbf{r}', \mathbf{r})\cdot\mathbf{e})]$$

$$= \int_{V'} d\tau'\, \mathbf{E}(\mathbf{r}')\cdot\mathbf{e}\,\delta(\mathbf{r}' - \mathbf{r}) = \mathbf{E}(\mathbf{r})\cdot\mathbf{e}$$

if the point $\mathbf{r}$ is contained within the region V'; otherwise the volume integral vanishes. Therefore the electric field within the region V' is related to the tangential components of the electric and magnetic fields on the surface S' by

$$\begin{aligned} \mathbf{E}(\mathbf{r}) = &-ik\int_{S'} dS'\, (\mathbf{n}' \times \mathbf{H}(\mathbf{r}'))\cdot\mathbf{\Gamma}^{(1)}(\mathbf{r}', \mathbf{r}) \\ &- \int_{S'} dS'\, (\mathbf{n}' \times \mathbf{E}(\mathbf{r}'))\cdot(\nabla' \times \mathbf{\Gamma}^{(1)}(\mathbf{r}', \mathbf{r})). \end{aligned} \tag{3.18}$$

The physical interpretation of this result can be made more apparent by employing the theorems (15), (16) to rewrite (18) in the form

$$\begin{aligned} \mathbf{E}(\mathbf{r}) = &-ik\int_{S'} \mathbf{\Gamma}^{(1)}(\mathbf{r}, \mathbf{r}')\cdot(\mathbf{n}' \times \mathbf{H}(\mathbf{r}'))\, dS' \\ &- \nabla \times \int_{S'} \mathbf{\Gamma}^{(2)}(\mathbf{r}, \mathbf{r}')\cdot(\mathbf{n}' \times \mathbf{E}(\mathbf{r}'))\, dS'. \end{aligned} \tag{3.19}$$

Recalling the definitions of the Green's functions in terms of the fields of point currents (equations (5) and (9)), we observe that (19) is just the electric field which would be produced by an electric surface current density

$$\mathbf{K}(\mathbf{r}) = -\frac{c}{4\pi}\mathbf{n} \times \mathbf{H}(\mathbf{r}), \tag{3.20}$$

and a magnetic surface current density

$$\mathbf{K}^*(\mathbf{r}) = \frac{c}{4\pi}\mathbf{n} \times \mathbf{E}(\mathbf{r}) \tag{3.21}$$

located on the surface S' (with outward normal $\mathbf{n}$). If, therefore, at the same time the actual field is removed, and the surface currents (20), (21) are caused to flow, the field within V' will remain unaltered, while the field outside V' will be reduced to zero.

When the surface S' coincides with S, the first integral in (18) vanishes, and we obtain an expression for $\mathbf{E}$ in terms of the tangential electric field alone, viz:

$$\mathbf{E}(\mathbf{r}) = -\int_S dS'\,(\mathbf{n}' \times \mathbf{E}(\mathbf{r}'))\cdot(\nabla' \times \mathbf{\Gamma}^{(1)}(\mathbf{r}', \mathbf{r})) \tag{3.22}$$

which explicitly demonstrates the first part of the existence theorem that motivates this discussion. It may now be remarked that, inasmuch as (18) implies no particular boundary conditions on $\mathbf{\Gamma}$, we may write the alternative expression

$$\begin{aligned} \mathbf{E}(\mathbf{r}) = &-ik\int_{S'} dS'\,(\mathbf{n}' \times \mathbf{H}(\mathbf{r}'))\cdot\mathbf{\Gamma}^{(2)}(\mathbf{r}', \mathbf{r}) \\ &-\int_{S'} dS'\,(\mathbf{n}' \times \mathbf{E}(\mathbf{r}'))\cdot(\nabla' \times \mathbf{\Gamma}^{(2)}(\mathbf{r}', \mathbf{r})), \end{aligned} \tag{3.23}$$

which, when S' coincides with S, reduces to

$$\mathbf{E}(\mathbf{r}) = -ik\int_S dS'\,(\mathbf{n}' \times \mathbf{H}(\mathbf{r}'))\cdot\mathbf{\Gamma}^{(2)}(\mathbf{r}', \mathbf{r}) \tag{3.24}$$

the explicit formulation of the second statement of the fundamental existence theorem.

The magnetic field can be calculated directly from the expression (19) for the electric field, or, more elegantly, by repeating the steps which led to (18), but replacing $\mathbf{E}$ by $\mathbf{H}$ and $\mathbf{\Gamma}^{(1)}$ by $\mathbf{\Gamma}^{(2)}$. It is evident that all that is necessary is to perform this substitution in the final formula, provided one also replaces $\mathbf{H}$ by $-\mathbf{E}$ (the minus sign is required to preserve the Maxwell equations). Hence

$$\begin{aligned} \mathbf{H}(\mathbf{r}) = &\;ik\int_{S'} dS'\,(\mathbf{n}' \times \mathbf{E}(\mathbf{r}'))\cdot\mathbf{\Gamma}^{(2)}(\mathbf{r}', \mathbf{r}) \\ &-\int_{S'} dS'\,(\mathbf{n}' \times \mathbf{H}(\mathbf{r}'))\cdot(\nabla' \times \mathbf{\Gamma}^{(2)}(\mathbf{r}', \mathbf{r})), \end{aligned} \tag{3.25}$$

or equivalently,

$$\begin{aligned} \mathbf{H}(\mathbf{r}) = &\;ik\int_{S'} \mathbf{\Gamma}^{(2)}(\mathbf{r}, \mathbf{r}')\cdot(\mathbf{n}' \times \mathbf{E}(\mathbf{r}'))\,dS' \\ &-\nabla \times \int_{S'} \mathbf{\Gamma}^{(1)}(\mathbf{r}, \mathbf{r}')\cdot(\mathbf{n}' \times \mathbf{H}(\mathbf{r}'))\,dS'. \end{aligned} \tag{3.26}$$

The magnetic field obtained from (19) is easily seen to agree with (26). If the surface S' coincides with S,

$$\mathbf{H}(\mathbf{r}) = ik \int_S \mathbf{\Gamma}^{(2)}(\mathbf{r}, \mathbf{r}')\cdot(\mathbf{n}' \times \mathbf{E}(\mathbf{r}'))\, dS', \tag{3.27}$$

indicating that the field can be produced by the presence of suitable magnetic currents on the surface S. Replacing $\mathbf{\Gamma}^{(2)}$ by $\mathbf{\Gamma}^{(1)}$ in (25), we have the alternative expression

$$\begin{aligned} \mathbf{H}(\mathbf{r}) = ik \int_{S'} dS'\, (\mathbf{n}' \times \mathbf{E}(\mathbf{r}'))\cdot\mathbf{\Gamma}^{(1)}(\mathbf{r}', \mathbf{r}) \\ - \int_{S'} dS'\, (\mathbf{n}' \times \mathbf{H}(\mathbf{r}'))\cdot(\nabla' \times \mathbf{\Gamma}^{(1)}(\mathbf{r}', \mathbf{r})), \end{aligned} \tag{3.28}$$

which, when S' coincides with S, reduces to

$$\mathbf{H}(\mathbf{r}) = -\int_S dS'\, (\mathbf{n}' \times \mathbf{H}(\mathbf{r}'))\cdot(\nabla' \times \mathbf{\Gamma}^{(1)}(\mathbf{r}', \mathbf{r})). \tag{3.29}$$

Equations (27) and (29) are particular manifestations of the fundamental existence theorem.

In concluding this section, we remark on the free space fields generated by an electric current on the surface of a perfect conductor. Since the current density is related to the tangential magnetic field at the surface by (20), where the normal $\mathbf{n}'$ points into the conductor, we have

$$\mathbf{E}(\mathbf{r}) = -ik \int_S dS'\, (\mathbf{n}' \times \mathbf{H}(\mathbf{r}'))\cdot\mathbf{\Gamma}^{(0)}(\mathbf{r}', \mathbf{r}), \tag{3.30}$$

and

$$\mathbf{H}(\mathbf{r}) = -\int_S dS'\, (\mathbf{n}' \times \mathbf{H}(\mathbf{r}'))\cdot(\nabla' \times \mathbf{\Gamma}^{(0)}(\mathbf{r}', \mathbf{r})). \tag{3.31}$$

For a plane current sheet, the current density is the difference in tangential component of the magnetic fields on opposite sides of the sheet.

4. *First Variational Principle*

With the stock of information concerning tensor Green's functions, we resume consideration of the diffraction problem. In this section, the development is based on the existence theorem as it relates to boundary values of the tangential electric field. Thus, in the half space bounded by the plane $z = 0$ and an infinitely remote surface where z is positive, we obtain from (3.22) and (3.27),

$$
\begin{aligned}
\mathbf{E}_+(\mathbf{r}) &= \int_{S_1} (\mathbf{e}_z \times \mathbf{E}(\varrho'))\cdot(\nabla' \times \mathbf{\Gamma}_+^{(1)}(x', y', 0, \mathbf{r}))\, dS' \\
\mathbf{H}_+(\mathbf{r}) &= -ik \int_{S_1} \mathbf{\Gamma}_+^{(2)}(\mathbf{r}, \varrho')\cdot(\mathbf{e}_z \times \mathbf{E}(\varrho'))\, dS'
\end{aligned}
\qquad z \geq 0 \tag{4.1}
$$

where ϱ denotes a position vector in the plane of the screen, and $\mathbf{\Gamma}_+^{(1),(2)}$ are the half space electric and magnetic field tensor Green's functions. The integrals in (1) extend over the aperture only; on the remainder of the plane $z = 0$, the surface integrals vanish by virtue of the boundary condition (2.3) for the tangential electric field. Moreover, the surface at infinity does not contribute, as can be inferred from the known (radial) behavior of the electric field in the wave zone and the asymptotic properties of the Green's functions.

On the other side of the screen,

$$
\begin{aligned}
\mathbf{E}_-(\mathbf{r}) &= \mathbf{E}_0(\mathbf{r}) - \int_{S_1} (\mathbf{e}_z \times \mathbf{E}(\varrho'))\cdot(\nabla' \times \mathbf{\Gamma}_-^{(1)}(x', y', 0, \mathbf{r}))\, dS' \\
\mathbf{H}_-(\mathbf{r}) &= \mathbf{H}_0(\mathbf{r}) + ik \int_{S_1} \mathbf{\Gamma}_-^{(2)}(\mathbf{r}, \varrho')\cdot(\mathbf{e}_z \times \mathbf{E}(\varrho'))\, dS'
\end{aligned}
\qquad z \leq 0 \tag{4.2}
$$

in view of a change in the sense of the positive normal at the plane $z = 0$; $\mathbf{E}_0$, $\mathbf{H}_0$ are given explicitly in (2.5), and

$$\mathbf{\Gamma}_-(\mathbf{r}, \mathbf{r}') = \mathbf{\Gamma}_+(\mathbf{r} - 2\mathbf{e}_z\cdot\mathbf{r}\mathbf{e}_z, \mathbf{r}' - 2\mathbf{e}_z\cdot\mathbf{r}'\mathbf{e}_z).$$

Combining (2.1) and (4.1) in accordance with (2.9) we obtain

$$
\begin{aligned}
&\mathbf{e}_z \times \mathbf{h}' \exp\{ik\mathbf{n}'\cdot\varrho\} \\
&\qquad = -ik\mathbf{e}_z \times \int_{S_1} \mathbf{\Gamma}_+^{(2)}(\varrho, \varrho')\cdot(\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\varrho'))\, dS' \\
&\qquad = -2ik\mathbf{e}_z \times \int_{S_1} \mathbf{\Gamma}^{(0)}(\varrho, \varrho')\cdot(\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\varrho'))\, dS', \qquad \varrho \text{ in } S_1
\end{aligned}
\tag{4.3}
$$

as a vector integral equation to specify the tangential electric aperture field, the latter being explicitly linked with the propagation direction of the incident plane wave. Were the solution of (3) generally feasible, a single integration according to (1) or (2) would determine the free space fields at any point. To alleviate the practical difficulties of such a program, we shall devise an approximation procedure for calculating the fields at large distance from the aperture.

For the distant transmitted fields (1), we employ the asymptotic forms of the Green's functions (cf. (3.12), (3.13))

$$\Gamma_+^{(1),(2)}(\mathbf{r}, \mathbf{r}')$$

$$\simeq \frac{1}{4\pi}\left(\boldsymbol{\varepsilon}\,\frac{\exp\{ik(r - \mathbf{n}\cdot\mathbf{r}')\}}{r} \mp (\boldsymbol{\varepsilon} - 2\mathbf{e}_z\mathbf{e}_z)\,\frac{\exp\{ik(r - \mathbf{n}\cdot(\mathbf{r}' - 2\mathbf{e}_z\mathbf{e}_z\cdot\mathbf{r}'))\}}{r}\right)$$

$$(4.4)\qquad - \frac{1}{4\pi k^2}\nabla\nabla'\left(\frac{\exp\{ik(r - \mathbf{n}\cdot\mathbf{r}')\}}{r} \mp \frac{\exp\{ik(r - \mathbf{n}\cdot(\mathbf{r}' - 2\mathbf{e}_z\mathbf{e}_z\cdot\mathbf{r}'))\}}{r}\right),$$

$$\mathbf{n} = \frac{\mathbf{r}}{r}, \qquad r \to \infty .$$

Thus, with elementary vector manipulation,

$$\mathbf{E}_+(\mathbf{r}) \simeq \int_{S_1} (\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho}')) \times \nabla'\cdot \frac{1}{4\pi}\left[\boldsymbol{\varepsilon}\,\frac{\exp\{ik(r - \mathbf{n}\cdot\mathbf{r}')\}}{r}\right.$$

$$(4.5)\qquad \left. - (\boldsymbol{\varepsilon} - 2\mathbf{e}_z\mathbf{e}_z)\,\frac{\exp\{ik(r - \mathbf{n}\cdot(\mathbf{r}' - 2\mathbf{e}_z\mathbf{e}_z\cdot\mathbf{r}'))\}}{r}\right] dS'$$

$$= -\frac{ik}{4\pi}\frac{e^{ikr}}{r}\int_{S_1} (\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho})) \times [\mathbf{n}\cdot\boldsymbol{\varepsilon} - (\mathbf{n} - 2\mathbf{e}_z\mathbf{e}_z\cdot\mathbf{n})$$

$$\cdot\,(\boldsymbol{\varepsilon} - 2\mathbf{e}_z\mathbf{e}_z)]\exp\{-ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\, dS.$$

To simplify (5), we observe that, using the abbreviation $\mathbf{V} = \mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}$,

$$\mathbf{V} \times \mathbf{n}\cdot\boldsymbol{\varepsilon} - \mathbf{V} \times (\mathbf{n} - 2\mathbf{e}_z\mathbf{e}_z\cdot\mathbf{n})\cdot(\boldsymbol{\varepsilon} - 2\mathbf{e}_z\mathbf{e}_z)$$

$$= 2\mathbf{e}_z(\mathbf{V} \times \mathbf{n}\cdot\mathbf{e}_z) + 2(\mathbf{e}_z\cdot\mathbf{n})\mathbf{V} \times \mathbf{e}_z = -2\mathbf{n} \times (\mathbf{e}_z \times (\mathbf{V} \times \mathbf{e}_z))$$

$$= -2\mathbf{n} \times \mathbf{V} + 2(\mathbf{e}_z\cdot\mathbf{V})\mathbf{n} \times \mathbf{e}_z = -2\mathbf{n} \times \mathbf{V},$$

since $\mathbf{e}_z\cdot\mathbf{V} = 0$. Hence,

$$(4.6)\qquad \mathbf{E}_+(\mathbf{r}) \simeq \mathbf{n} \times \mathbf{A}(\mathbf{n}, \mathbf{n}')\,\frac{e^{ikr}}{r}, \qquad r \to \infty$$

where

$$(4.7)\qquad \mathbf{A}(\mathbf{n}, \mathbf{n}') = \frac{ik}{2\pi}\int_{S_1} \mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho})\exp\{-ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\, dS.$$

Further,

$$\mathbf{H}_+(\mathbf{r}) \simeq -\frac{ik}{4\pi}\int_{S_1}\left[\boldsymbol{\varepsilon}\,\frac{\exp\{ik(r - \mathbf{n}\cdot\boldsymbol{\varrho}')\}}{r} + (\boldsymbol{\varepsilon} - 2\mathbf{e}_z\mathbf{e}_z)\,\frac{\exp\{ik(r - \mathbf{n}\cdot\boldsymbol{\varrho}')\}}{r}\right]$$

$$\cdot\,(\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho}'))\, dS'$$

$$+ \frac{i}{4\pi k}\int_{S_1} \nabla\nabla'\left[\frac{\exp\{ik(r - \mathbf{n}\cdot\mathbf{r}')\}}{r}\right.$$

$$\left. - \frac{\exp\{ik(r - \mathbf{n}\cdot(\mathbf{r}' - 2\mathbf{e}_z\mathbf{e}_z\cdot\mathbf{r}'))\}}{r}\right]_{z'=0} \cdot(\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho}'))\, dS'$$

$$= \frac{ik}{2\pi}\frac{e^{ikr}}{r}\int_{S_1} [\mathbf{nn}\cdot(\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho})) - (\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho}))] \exp\{-ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\, dS,$$

and thus

$$\mathbf{H}_+(\mathbf{r}) \simeq \mathbf{n} \times (\mathbf{n} \times \mathbf{A}(\mathbf{n}, \mathbf{n}'))\frac{e^{ikr}}{r}, \qquad r \to \infty. \tag{4.8}$$

The results (6) and (8) show that electric and magnetic fields in the wave zone ($kr \gg 1$) are of equal magnitude and together with the propagation vector form a mutually orthogonal triad.

Let us next perform scalar premultiplication with $\mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho})$ in the integral equation (3), and integrate over the aperture domain. There results

$$\int_{S_1} \mathbf{e}_z \times \mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho})\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}')\cdot\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho}')\, dS\, dS'$$

$$= \frac{i}{2k}\mathbf{h}'\cdot\int_{S_1} \mathbf{e}_z \times \mathbf{E}_{\mathbf{n}''}(\boldsymbol{\varrho}) \exp\{ik\mathbf{n}'\cdot\boldsymbol{\varrho}\}\, dS = \frac{\pi}{k^2}\mathbf{h}'\cdot\mathbf{A}(-\mathbf{n}', \mathbf{n}'') \tag{4.9}$$

$$= \frac{i}{2k}\mathbf{h}''\cdot\int_{S_1} \mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho}) \exp\{ik\mathbf{n}''\cdot\boldsymbol{\varrho}\}\, dS = \frac{\pi}{k^2}\mathbf{h}''\cdot\mathbf{A}(-\mathbf{n}'',\mathbf{n}'),$$

making use of evident symmetry in the first term as regards exchange of the indices $\mathbf{n}'$, $\mathbf{n}''$. From (9), we learn that

$$\frac{ik}{2\pi}\left(\mathbf{h}''\cdot\int_{S_1} \mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho}) \exp\{-ik\mathbf{n}''\cdot\boldsymbol{\varrho}\}\, dS + \mathbf{h}'\cdot\int_{S_1} \mathbf{e}_z \times \mathbf{E}_{-\mathbf{n}''}(\boldsymbol{\varrho})\right.$$

$$\left.\cdot \exp\{ik\mathbf{n}'\cdot\boldsymbol{\varrho}\}\, dS\right) \tag{4.10}$$

$$- \frac{k^2}{\pi}\int_{S_1} \mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho})\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}')\cdot\mathbf{e}_z \times \mathbf{E}_{-\mathbf{n}''}(\boldsymbol{\varrho}')\, dS\, dS'$$

$$= \mathbf{h}''\cdot\mathbf{A}(\mathbf{n}'', \mathbf{n}') = \mathbf{h}'\cdot\mathbf{A}(-\mathbf{n}', -\mathbf{n}''),$$

and this expression for $\mathbf{h}''\cdot\mathbf{A}(\mathbf{n}'', \mathbf{n}')$, or $\mathbf{h}'\cdot\mathbf{A}(-\mathbf{n}', -\mathbf{n}'')$, is stationary with respect to independent variations of $\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}$, $\mathbf{e}_z \times \mathbf{E}_{-\mathbf{n}''}$ relative to tangential electric aperture fields specified by integral equations of the form (3). On

carrying out individual scale transformations of the aperture fields in (10), we arrive at the useful stationary homogeneous forms

$$\frac{1}{\mathbf{h}''\cdot\mathbf{A}(\mathbf{n}'',\mathbf{n}')} = \frac{1}{\mathbf{h}'\cdot\mathbf{A}(-\mathbf{n}',-\mathbf{n}'')} = -4\pi \tag{4.11}$$
$$\cdot\frac{\int_{S_1}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}'}(\varrho)\cdot\mathbf{\Gamma}^{(0)}(\varrho,\varrho')\cdot\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}''}(\varrho')dS\,dS'}{(\mathbf{h}''\cdot\int_{S_1}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}'}(\varrho)\exp\{-ik\mathbf{n}''\cdot\varrho\}dS)(\mathbf{h}'\cdot\int_{S_1}\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}''}(\varrho)\exp\{ik\mathbf{n}'\cdot\varrho\}dS)}.$$

Since the integral equation (3) reveals that the projection of $\mathbf{n}'$ on the plane of the screen characterizes the aperture field $\mathbf{e}_z \times \mathbf{E}_{\mathbf{n}'}$, when excitation is restricted to the half space $z < 0$, a reversal in the sign of $\mathbf{n}'$ implies an increase of π in the azimuthal angle φ', the magnetic polarization vector being unchanged. However, as the aperture presents the same aspect in either half space, the field $\mathbf{e}_z \times \mathbf{E}_{-\mathbf{n}'}$ can be ascribed to plane wave excitation in the direction opposite to $\mathbf{n}'$, retaining the original magnetic polarization and reversing the electric polarization to ensure the correct sense of propagation. The equality of $\mathbf{h}''\cdot\mathbf{A}(\mathbf{n}'', \mathbf{n}')$ and $\mathbf{h}'\cdot\mathbf{A}(-\mathbf{n}', -\mathbf{n}'')$ is thus in the nature of a reciprocity relation for the diffracted amplitudes which accompany excitation and observation along a pair of directions in space.

5. *Second Variational Principle*

An independent variational principle stems from the existence theorem which relates to boundary values of the tangential magnetic field. For the half space $z > 0$ with boundary values ($z = 0$)

$$\begin{aligned}\mathbf{K}^+(\varrho) &= \mathbf{e}_z \times \mathbf{H}_+(\varrho), \qquad \varrho \text{ on } S_2\\ &= \mathbf{e}_z \times \mathbf{H}^{\mathrm{inc}}(\varrho), \qquad \varrho \text{ in } S_1\end{aligned} \tag{5.1}$$

the last from (2.9), we have according to (3.24) and (3.29),

$$\begin{aligned}\mathbf{E}_+(\mathbf{r}) &= ik\int_{S_2}\mathbf{\Gamma}_+^{(2)}(\mathbf{r},\varrho')\cdot\mathbf{K}_{\mathbf{n}'}^+(\varrho')\,dS'\\ &\qquad + ik\int_{S_1}\mathbf{\Gamma}_+^{(2)}(\mathbf{r},\varrho')\cdot(\mathbf{e}_z\times\mathbf{h}').\exp\{ik\mathbf{n}'\cdot\varrho'\}\,dS'\\ \mathbf{H}_+(\mathbf{r}) &= \int_{S_2}\mathbf{K}_{\mathbf{n}'}^+(\varrho')\cdot\nabla'\times\mathbf{\Gamma}_+^{(1)}(x',y',0,\mathbf{r})\,dS'\\ &\qquad + \int_{S_1}(\mathbf{e}_z\times\mathbf{h}')\exp\{ik\mathbf{n}'\cdot\varrho'\}\cdot\nabla'\times\mathbf{\Gamma}_+^{(1)}(x',y',0,\mathbf{r})\,dS'.\end{aligned} \tag{5.2}$$

The function $\mathbf{K}^+$, multiplied by $c/4\pi$ to obtain the equivalent surface current density, tends to zero with increasing distance from the aperture.

Likewise, the boundary distribution

(5.3)
$$\begin{aligned} \mathbf{K}^-(\varrho) &= \mathbf{e}_z \times \mathbf{H}_-(\varrho), \qquad \varrho \text{ on } S_2 \\ &= \mathbf{e}_z \times \mathbf{H}^{\mathrm{inc}}(\varrho), \qquad \varrho \text{ in } S_1 \end{aligned}$$

for the half space $z < 0$, yields

(5.4)
$$\begin{aligned} \mathbf{E}_-(\mathbf{r}) = {} & \mathbf{e}' \exp\{ik\mathbf{n}'\cdot\mathbf{r}\} + \mathbf{e}'\cdot(\boldsymbol{\varepsilon} - 2\mathbf{e}_z\mathbf{e}_z) \exp\{ik\mathbf{n}'\cdot(\mathbf{r} - 2\mathbf{e}_z\mathbf{e}_z\cdot\mathbf{r})\} \\ & - ik \int_{S_2} \boldsymbol{\Gamma}_-^{(2)}(\mathbf{r}, \varrho')\cdot\mathbf{K}_{\mathbf{n}'}^-(\varrho')\, dS' \\ & - ik \int_{S_1} \boldsymbol{\Gamma}_-^{(2)}(\mathbf{r}, \varrho')\cdot(\mathbf{e}_z \times \mathbf{h}') \exp\{ik\mathbf{n}'\cdot\varrho'\}\, dS', \\ \mathbf{H}_-(\mathbf{r}) = {} & \mathbf{h}' \exp\{ik\mathbf{n}'\cdot\mathbf{r}\} - \mathbf{h}'\cdot(\boldsymbol{\varepsilon} - 2\mathbf{e}_z\mathbf{e}_z) \exp\{ik\mathbf{n}'\cdot(\mathbf{r} - 2\mathbf{e}_z\mathbf{e}_z\cdot\mathbf{r})\} \\ & - \int_{S_2} \mathbf{K}_{\mathbf{n}'}^-(\varrho')\cdot\nabla' \times \boldsymbol{\Gamma}_-^{(1)}(x', y', 0, \mathbf{r})\, dS' \\ & - \int_{S_1} (\mathbf{e}_z \times \mathbf{h}') \exp\{ik\mathbf{n}'\cdot\varrho'\}\cdot\nabla' \times \boldsymbol{\Gamma}_-^{(1)}(x', y', 0, \mathbf{r})\, dS' \end{aligned}$$

where the integrated terms are due to the infinitely remote part of the boundary surface. At large distance from the aperture, $\mathbf{K}^-$ (apart from the factor $c/4\pi$) becomes identical with the infinite screen current density,

$$\mathbf{K}_0(\varrho) = \mathbf{e}_z \times \mathbf{H}_0(\varrho) = 2\mathbf{e}_z \times \mathbf{H}^{\mathrm{inc}}(\varrho), \qquad \varrho \text{ on } S_1 + S_2 . \tag{5.5}$$

From the requirement of vanishing tangential electric field on the shadow face of the screen, it follows that

(5.6)
$$\begin{aligned} & \mathbf{e}_z \times \int_{S_2} \boldsymbol{\Gamma}^{(0)}(\varrho, \varrho')\cdot\mathbf{K}_{\mathbf{n}'}^+(\varrho')\, dS' \\ & = -\mathbf{e}_z \times \int_{S_1} \boldsymbol{\Gamma}^{(0)}(\varrho, \varrho')\cdot(\mathbf{e}_z \times \mathbf{h}') \exp\{ik\mathbf{n}'\cdot\varrho'\}\, dS', \qquad \varrho \text{ on } S_2 \end{aligned}$$

since only tangential components of $\boldsymbol{\Gamma}_+^{(2)}$ are involved. Similarly, on the illuminated face of the screen,

(5.7)
$$\begin{aligned} \mathbf{e}_z \times \mathbf{e}' \exp\{ik\mathbf{n}'\cdot\varrho\} = {} & ik\mathbf{e}_z \times \int_{S_2} \boldsymbol{\Gamma}^{(0)}(\varrho, \varrho')\cdot\mathbf{K}_{\mathbf{n}'}^-(\varrho')\, dS' \\ & + ik\mathbf{e}_z \times \int_{S_1} \boldsymbol{\Gamma}^{(0)}(\varrho, \varrho')\cdot(\mathbf{e}_z \times \mathbf{h}') \exp\{ik\mathbf{n}'\cdot\varrho'\}\, dS', \qquad \varrho \text{ on } S_2 . \end{aligned}$$

The integral equations (6) and (7) for $\mathbf{K}^+_{\mathbf{n}'}$ and $\mathbf{K}^-_{\mathbf{n}'}$, like that encountered in the previous section, have no ready solution. It may be noted that for a completely infinite screen, equation (7) becomes

$$\begin{aligned}\mathbf{e}_z \times \mathbf{e}' \exp\{ik\mathbf{n}'\cdot\boldsymbol{\varrho}\} &= 2ik\mathbf{e}_z \times \int_{S_1+S_2} \boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}') \\ &\quad \cdot(\mathbf{e}_z \times \mathbf{h}') \exp\{ik\mathbf{n}'\cdot\boldsymbol{\varrho}'\}\, dS', \qquad \boldsymbol{\varrho} \text{ on } S_1 + S_2\end{aligned} \tag{5.8}$$

a result which can be directly verified by use of (3.12) and the integral representation

$$\begin{aligned}\frac{\exp\{ik\,|\mathbf{r}-\mathbf{r}'|\}}{4\pi\,|\mathbf{r}-\mathbf{r}'|} &= \frac{i}{8\pi^2}\int_{-\infty}^{\infty} \frac{dk_x\, dk_y}{(k^2 - k_x^2 - k_y^2)^{1/2}} \\ &\quad \cdot \exp\{i(k_x(x-x') + k_y(y-y') + (k^2-k_x^2-k_y^2)^{1/2}\,|z-z'|)\}.\end{aligned} \tag{5.9}$$

Using (4.4) we deduce that the transmitted fields remote from the aperture are

$$\left.\begin{aligned}\mathbf{E}_+(\mathbf{r}) &\simeq -\mathbf{n} \times (\mathbf{n} \times \mathbf{B}(\mathbf{n}, \mathbf{n}'))\,\frac{e^{ikr}}{r} \\ \mathbf{H}_+(\mathbf{r}) &\simeq \mathbf{n} \times \mathbf{B}(\mathbf{n}, \mathbf{n}')\,\frac{e^{ikr}}{r}\end{aligned}\right\} \qquad r \to \infty \tag{5.10}$$

where

$$\begin{aligned}\mathbf{B}(\mathbf{n}, \mathbf{n}') = \frac{ik}{2\pi}\Bigg[&\int_{S_2} \mathbf{K}^+_{\mathbf{n}'}(\boldsymbol{\varrho}) \exp\{-ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\, dS \\ &+ (\mathbf{e}_z \times \mathbf{h}') \int_{S_1} \exp\{ik(\mathbf{n}'-\mathbf{n})\cdot\boldsymbol{\varrho}\}\, dS\Bigg].\end{aligned} \tag{5.11}$$

In consequence of the integral equation for $\mathbf{K}^+_{\mathbf{n}'}$,

$$\begin{aligned}&\int_{S_2} \mathbf{K}^+_{\mathbf{n}''}(\boldsymbol{\varrho})\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}')\cdot\mathbf{K}^+_{\mathbf{n}'}(\boldsymbol{\varrho}')\, dS\, dS' \\ &= -\int_{S_1} dS' \int_{S_2} dS\, \mathbf{K}^+_{\mathbf{n}''}(\boldsymbol{\varrho})\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}')\cdot(\mathbf{e}_z \times \mathbf{h}') \exp\{ik\mathbf{n}'\cdot\boldsymbol{\varrho}'\} \\ &= -\int_{S_1} dS' \int_{S_2} dS\, \mathbf{K}^+_{\mathbf{n}'}(\boldsymbol{\varrho})\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}')\cdot(\mathbf{e}_z \times \mathbf{h}'') \exp\{ik\mathbf{n}''\cdot\boldsymbol{\varrho}'\},\end{aligned}$$

and furthermore, by appealing to (6) and (8),

$$\int_{S_1} dS' \int_{S_2} dS\, K^{+}_{n''}(\varrho)\cdot \Gamma^{(0)}(\varrho, \varrho')\cdot(e_z \times h') \exp\{ik n'\cdot\varrho'\}$$

$$= \int_{S_1+S_2} dS' \int_{S_2} dS\, \mathbf{K}^{+}_{n''}(\varrho)\cdot \Gamma^{(0)}(\varrho, \varrho')\cdot(e_z \times h') \exp\{ik n'\cdot\varrho'\}$$

$$- \int_{S_2} dS'\, dS\, \mathbf{K}^{+}_{n''}(\varrho)\cdot \Gamma^{(0)}(\varrho, \varrho')\cdot(e_z \times h') \exp\{ik n'\cdot\varrho'\}$$

$$= \int_{S_1+S_2} dS' \int_{S_2} dS\, \mathbf{K}^{+}_{n''}(\varrho)\cdot \Gamma^{(0)}(\varrho, \varrho')\cdot(e_z \times h') \exp\{ik n'\cdot\varrho'\}$$

$$+ \int_{S_1+S_2} dS' \int_{S_1} dS\, (e_z \times h'') \exp\{ik n''\cdot\varrho\}\cdot \Gamma^{(0)}(\varrho, \varrho')\cdot(e_z \times h')$$

$$\cdot \exp\{ik n'\cdot\varrho'\}$$

$$- \int_{S_1} dS'\, dS\, (e_z \times h'') \exp\{ik n''\cdot\varrho\}\Gamma^{(0)}(\varrho, \varrho')\cdot(e_z \times h') \exp\{ik n'\cdot\varrho'\}$$

$$= \frac{1}{2ik} e'\cdot\left[\int_{S_2} \mathbf{K}^{+}_{n''}(\varrho) \exp\{ik n'\cdot\varrho\}\, dS\right.$$

$$\left. + \int_{S_1} (e_z \times h'') \exp\{ik(n' + n'')\cdot\varrho\}\, dS\right]$$

$$- \int_{S_1} dS\, dS'\, (e_z \times h') \exp\{ik n'\cdot\varrho\}\cdot \Gamma^{(0)}(\varrho, \varrho')\cdot(e_z \times h'') \exp\{ik n''\cdot\varrho'\}.$$

Thus

$$\int_{S_2} \mathbf{K}^{+}_{n'}(\varrho)\cdot \Gamma^{(0)}(\varrho, \varrho')\cdot \mathbf{K}^{+}_{-n''}(\varrho')\, dS\, dS'$$

$$+ \int_{S_1} dS' \int_{S_2} dS\, \mathbf{K}^{+}_{n'}(\varrho)\cdot \Gamma^{(0)}(\varrho, \varrho')\cdot(e_z \times h'') \exp\{-ik n''\cdot\varrho'\}$$

$$+ \int_{S_1} dS' \int_{S_2} dS\, \mathbf{K}^{+}_{-n''}(\varrho)\cdot \Gamma^{(0)}(\varrho, \varrho')\cdot(e_z \times h') \exp\{ik n'\cdot\varrho'\} \tag{5.12}$$

$$= -\frac{\pi}{k^2} e''\cdot[\mathbf{B}(n'', n') - \mathbf{B}_K(n'', n')]$$

$$= -\frac{\pi}{k^2} e'\cdot[\mathbf{B}(-n', -n'') - \mathbf{B}_K(-n', -n'')],$$

where

$$\mathbf{e}''\cdot\mathbf{B}_K(\mathbf{n}'', \mathbf{n}') = \mathbf{e}'\cdot\mathbf{B}_K(-\mathbf{n}', -\mathbf{n}'')$$

$$(5.13)\qquad = -\frac{k^2}{\pi}\int_{S_1} (\mathbf{e}_z \times \mathbf{h}') \exp\{ik\mathbf{n}'\cdot\boldsymbol{\varrho}\}\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}')\cdot(\mathbf{e}_z \times \mathbf{h}'') \cdot \exp\{-ik\mathbf{n}''\cdot\boldsymbol{\varrho}'\}\, dS\, dS'.$$

In accord with considerations of the preceding section, the function $\mathbf{K}^+_{-\mathbf{n}'}$ arises from plane wave excitation along the direction obtained by rotating $\mathbf{n}'$ about the z-axis through an angle π. If $\mathbf{K}^+$ vanishes on the screen, the amplitudes $\mathbf{B}$, $\mathbf{B}_K$ are equal, and hence the latter befits a Kirchhoff approximation. It may be noted that the real part of $\mathbf{B}_K$ is a divergent integral, although compensating integrals of (12) render $\mathbf{B}$ entirely convergent. The expression (12) for $\mathbf{e}''\cdot\mathbf{B}(\mathbf{n}'', \mathbf{n}')$ or $\mathbf{e}'\cdot\mathbf{B}(-\mathbf{n}', -\mathbf{n}'')$ has the property that it is stationary with respect to independent variations of $\mathbf{K}^+_{\mathbf{n}'}$, $\mathbf{K}^+_{-\mathbf{n}''}$ relative to solutions of the integral equation (6).

The foregoing development can be modified slightly, so as to involve the screen current density

$$(5.14)\qquad \mathbf{K}(\boldsymbol{\varrho}) = \mathbf{K}^-(\boldsymbol{\varrho}) - \mathbf{K}^+(\boldsymbol{\varrho}), \qquad \boldsymbol{\varrho} \text{ on } S_2$$

instead of $\mathbf{K}^+$, $\mathbf{K}^-$ individually; this transformation reflects a change in geometrical basis, with the screen now regarded as an obstacle imbedded in free space. Following the discussion at the close of section III, and with particular reference to (3.30) and (3.31), the fields at any point are

$$(5.15)\qquad \begin{aligned} \mathbf{E}(\mathbf{r}) &= \mathbf{e}' \exp\{ik\mathbf{n}'\cdot\mathbf{r}\} - ik\int_{S_2} \mathbf{K}_{\mathbf{n}'}(\boldsymbol{\varrho}')\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho}', \mathbf{r})\, dS' \\ \mathbf{H}(\mathbf{r}) &= \mathbf{h}' \exp\{ik\mathbf{n}'\cdot\mathbf{r}\} - \int_{S_2} \mathbf{K}_{\mathbf{n}'}(\boldsymbol{\varrho}')\cdot\nabla' \times \boldsymbol{\Gamma}^{(0)}(x', y', 0, \mathbf{r})\, dS', \end{aligned}$$

and clearly, from (15) or (6) and (7), the integral equation to determine $\mathbf{K}$ is

$$(5.16)\qquad \mathbf{e}_z \times \mathbf{e}' \exp\{ik\mathbf{n}'\cdot\boldsymbol{\varrho}\} = ik\mathbf{e}_z \times \int_{S_2} \boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}')\cdot\mathbf{K}_{\mathbf{n}'}(\boldsymbol{\varrho}')\, dS', \quad \boldsymbol{\varrho} \text{ on } S_2 .$$

If the current density $\mathbf{K}$ is resolved into two parts,

$$(5.17)\qquad \mathbf{K}(\boldsymbol{\varrho}) = \overline{\mathbf{K}}(\boldsymbol{\varrho}) + \mathbf{K}_0(\boldsymbol{\varrho}), \qquad \boldsymbol{\varrho} \text{ on } S_2$$

with $\mathbf{K}_0$ given by (5), and hence $\overline{\mathbf{K}}$ a null function for $\varrho \to \infty$, it can be shown that

$$\begin{Bmatrix}\mathbf{E}_+(\mathbf{r})\\ \mathbf{E}_-(\mathbf{r})\end{Bmatrix} = \begin{Bmatrix}0\\ \mathbf{E}_0(\mathbf{r})\end{Bmatrix} + 2ik\int_{S_1} \mathbf{\Gamma}^{(0)}(\mathbf{r}, \boldsymbol{\varrho}')\cdot(\mathbf{e}_z \times \mathbf{h}') \exp\{ik\mathbf{n}'\cdot\boldsymbol{\varrho}'\}\, dS' - ik\int_{S_2} \mathbf{\Gamma}^{(0)}(\mathbf{r}, \boldsymbol{\varrho}')\cdot\overline{\mathbf{K}}_{\mathbf{n}'}(\boldsymbol{\varrho}')\, dS' \tag{5.18}$$

and

$$\begin{Bmatrix}\mathbf{H}_+(\mathbf{r})\\ \mathbf{H}_-(\mathbf{r})\end{Bmatrix} = \begin{Bmatrix}0\\ \mathbf{H}_0(\mathbf{r})\end{Bmatrix} + 2\int_{S_1} (\mathbf{e}_z \times \mathbf{h}') \exp\{ik\mathbf{n}'\cdot\boldsymbol{\varrho}'\}\cdot\nabla' \times \mathbf{\Gamma}^{(0)}(x', y', 0, \mathbf{r})\, dS' - \int_{S_2} \overline{\mathbf{K}}_{\mathbf{n}'}(\boldsymbol{\varrho}')\cdot\nabla' \times \mathbf{\Gamma}^{(0)}(x', y', 0, \mathbf{r})\, dS', \tag{5.19}$$

whence

$$\mathbf{e}_z \times \int_{S_2} \mathbf{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}')\cdot\overline{\mathbf{K}}_{\mathbf{n}'}(\boldsymbol{\varrho}')\, dS' = 2\mathbf{e}_z \times \int_{S_1} \mathbf{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}')\cdot(\mathbf{e}_z \times \mathbf{h}') \exp\{ik\mathbf{n}'\cdot\boldsymbol{\varrho}'\}\, dS', \qquad \boldsymbol{\varrho} \text{ on } S_2\,. \tag{5.20}$$

An inspection of (6) and (20) reveals

$$\overline{\mathbf{K}}(\boldsymbol{\varrho}) = -2\mathbf{K}^+(\boldsymbol{\varrho}) \tag{5.21}$$

as can be inferred also from the general considerations of section 2. It is evident that to obtain a stationary expression for $\mathbf{B}$ in terms of $\overline{\mathbf{K}}$, one need only perform the substitution (21) in (11) and (12). The Kirchoff approximation $\mathbf{B}_K$ is thus based on the induced current $\mathbf{K}_0$, unmodified by the presence of an aperture.

6. *Transmission Cross Section*

To extend the practicality of the variational principles, we develop their connection with a quantity of physical interest, the plane wave transmission cross section of the aperture. A suitable expression for the cross section, or ratio of transmitted energy flux to incident energy flux per unit area, is derived from the real part of the Poynting vector theorem in a non-dissipative source free region, (the asterisk denotes complex conjugate)

$$\mathcal{R}e \int \nabla \cdot \frac{c}{8\pi}\mathbf{E} \times \mathbf{H}^*\, d\tau = 0 \tag{6.1}$$

which involves the quantity

$$\mathbf{S} = \frac{c}{8\pi}\mathbf{E} \times \mathbf{H}^* \tag{6.2}$$

the real part of which is the average energy flux vector. On integrating (1) throughout the shadow half space, we establish a connection between the total energy flux at infinity and that through the aperture,

$$
\begin{aligned}
&\mathcal{R}e\,\frac{c}{8\pi}\int_{S_\infty} \mathbf{n}\cdot\mathbf{E}_+ \times \mathbf{H}_+^*\,dS \\
(6.3)\qquad &= \mathcal{R}e\,\frac{c}{8\pi}\int_{S_1} \mathbf{e}_z\cdot\mathbf{E}\times\mathbf{H}^*\,dS = \mathcal{R}e\,\frac{c}{8\pi}\int_{S_1} \mathbf{e}_z\times\mathbf{E}\cdot(\mathbf{H}^{\mathrm{inc}})^*\,dS,
\end{aligned}
$$

taking account of (2.3) and (2.9). Since the magnitude of (2) for the incident wave (2.1) is $c/8\pi$, the cross section becomes

$$
(6.4)\qquad \sigma(\mathbf{n}') = \mathcal{R}e\,\mathbf{h}'\cdot\int_{S_1} \mathbf{e}_z\times\mathbf{E}_{\mathbf{n}'}(\boldsymbol{\varrho})\exp\{-ik\mathbf{n}'\cdot\boldsymbol{\varrho}\}\,dS = \frac{2\pi}{k}\,\mathcal{I}m\,\mathbf{h}'\cdot\mathbf{A}(\mathbf{n}',\mathbf{n}'),
$$

and in the latter form, the stationary expression (4.11) for $\mathbf{h}'\cdot\mathbf{A}(\mathbf{n}',\mathbf{n}')$ is directly applicable.

It can be verified that

$$
(6.5)\qquad \mathcal{R}e\int_{S_1} \mathbf{e}_z\times\mathbf{E}\cdot(\mathbf{H}^{\mathrm{inc}})^*\,dS = -\mathcal{R}e\int_{S_1+S_2} \mathbf{e}_z\times\mathbf{H}\cdot(\mathbf{E}^{\mathrm{inc}})^*\,dS,
$$

which provides another form of the cross section

$$
\begin{aligned}
\sigma(\mathbf{n}') &= -\mathcal{R}e\,\mathbf{e}'\cdot\int_{S_1+S_2} \mathbf{e}_z\times\mathbf{H}_{\mathbf{n}'}(\boldsymbol{\varrho})\exp\{-ik\mathbf{n}'\cdot\boldsymbol{\varrho}\}\,dS \\
(6.6)\qquad &= -\frac{2\pi}{k}\,\mathcal{I}m\,\mathbf{e}'\cdot\mathbf{B}(\mathbf{n}',\mathbf{n}'),
\end{aligned}
$$

adapted to the stationary principle of (5.12). The same result is established also via the sequence of relations

$$
\mathbf{h}'\cdot\mathbf{A}(\mathbf{n}',\mathbf{n}') = -\mathbf{e}'\cdot\mathbf{n}'\times\mathbf{A}(\mathbf{n}',\mathbf{n}') = \mathbf{e}'\cdot\mathbf{n}'\times(\mathbf{n}'\times\mathbf{B}(\mathbf{n}',\mathbf{n}')) = -\mathbf{e}'\cdot\mathbf{B}(\mathbf{n}',\mathbf{n}'),
$$

since $\mathbf{e}'\cdot\mathbf{n}' = 0$.

Combining (4) and (6) with (4.11) and (5.12) respectively, we find

$$
\begin{aligned}
\sigma(\mathbf{n}) &= -\frac{1}{2k}\,\mathcal{I}m\Bigg[\mathbf{h}\cdot\textstyle\int_{S_1}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}}(\boldsymbol{\varrho})\exp\{-ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\,dS \\
(6.7)\qquad &\quad\cdot\frac{(\mathbf{h}\cdot\int_{S_1}\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}}(\boldsymbol{\varrho})\exp\{ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\,dS)}{\int_{S_1}\mathbf{e}_z\times\mathbf{E}_{\mathbf{n}}(\boldsymbol{\varrho})\cdot\Gamma^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\cdot\mathbf{e}_z\times\mathbf{E}_{-\mathbf{n}}(\boldsymbol{\varrho}')\,dS\,dS'}\Bigg]
\end{aligned}
$$

and

$$\sigma(\mathbf{n}) = \sigma_K(\mathbf{n}) - 2k\,\mathscr{I}m\left[\int_{S_1} dS' \int_{S_1} dS\,\mathbf{K}_{\mathbf{n}}^{+}(\varrho)\cdot\Gamma^{(0)}(\varrho,\varrho')\cdot(\mathbf{e}_z\times\mathbf{h})\exp\{-ik\mathbf{n}\cdot\varrho'\}\right.$$

(6.8)

$$\left.\cdot\frac{\int_{S_1} dS' \int_{S_1} dS\,\mathbf{K}_{-\mathbf{n}}^{+}(\varrho)\cdot\Gamma^{(0)}(\varrho,\varrho')\cdot(\mathbf{e}_z\times\mathbf{h})\exp\{ik\mathbf{n}\cdot\varrho'\}}{\int_{S_1}\mathbf{K}_{\mathbf{n}}^{+}(\varrho)\cdot\Gamma^{(0)}(\varrho,\varrho')\cdot\mathbf{K}_{-\mathbf{n}}^{+}(\varrho')\,dS\,dS'}\right]$$

where

$$\sigma_K(\mathbf{n}) = 2k\,\mathscr{I}m\int_{S_1}(\mathbf{e}_z\times\mathbf{h})\exp\{ik\mathbf{n}\cdot\varrho\}\cdot\mathbf{\Gamma}^{(0)}(\varrho,\varrho')\cdot(\mathbf{e}_z\times\mathbf{h})$$

(6.9)

$$\cdot\exp\{-ik\mathbf{n}\cdot\varrho'\}\,dS\,dS'$$

is the Kirchhoff contribution. In (7)-(9) we have a pair of stationary, homogeneous expressions suitable for evaluating the cross section. A comparison of their respective predictions, based on trial functions for $\mathbf{E}$ and $\mathbf{K}^{+}$, provides an estimate of the accuracy obtained, since the results are identical only if the correct functions are employed. Particular interest concerns features of the cross section at very long or short wave lengths compared with aperture dimensions; such details as frequency or angle dependence are available from the stationary principles without knowledge of the correct boundary field distributions, which merely fix the proper scale.

At long wave lengths, or low frequencies, the characteristics of diffraction by an aperture can be described quite generally, without restriction to a particular type of incident field. The secondary field is then attributed to a pair of electric and magnetic dipoles in the aperture. These dipoles are respectively normal and parallel to the plane of the aperture, and are related to the corresponding components of the fields $\mathbf{E}_0$, $\mathbf{H}_0$, which can be regarded as constants. The coefficients of proportionality, termed electric and magnetic polarizabilities, are independent of wave length and given in terms of aperture dimensions. For the electric dipole, a single polarizability occurs, whereas the magnetic dipole has an associated polarizability dyadic, as may be inferred from the relative orientations of moments and exciting fields. Symmetry of the dyadic implies that there are mutually perpendicular principal axes, along each of which the magnetic dipole moment is a multiple of the associated component of $\mathbf{H}_0$.

The tangential electric aperture field has separate contributions due to $\mathbf{E}_0$, $\mathbf{H}_0$, and for plane wave excitation takes the form (cf. Appendix 2)

(6.10) $$\mathbf{E}_t(\varrho) = \mathbf{e}_z\cdot\mathbf{e}\nabla\phi_1(\varrho) + ik(-\mathbf{h}\cdot\mathbf{l}\mathbf{m}\phi_2(\varrho) + \mathbf{h}\cdot\mathbf{m}\mathbf{l}\phi_3(\varrho)), \qquad \varrho \text{ in } S_1$$

where $\mathbf{l}$, $\mathbf{m}$ are unit vectors along the principal axes of the magnetic polarizability dyadic, viz.:

(6.11) $$\mathbf{e}_z\times\mathbf{l} = \mathbf{m}, \qquad \mathbf{e}_z\times\mathbf{m} = -\mathbf{l}, \qquad \mathbf{l}\cdot\mathbf{m} = 0.$$

The functions $\phi(\varrho)$ are individually real and frequency independent; a boundary

condition is necessary to ensure that the tangential component of $\mathbf{E}_t$ vanishes at the rim of the screen, where $\phi_1(\varrho)$ itself is zero.

If the electric field of an incident plane wave is perpendicular to the plane of incidence defined by $\mathbf{e}_z$ and $\mathbf{n}$, whence $\mathbf{e}_z \cdot \mathbf{e} = 0$, we find on use of (10), (11) and the expansion

$$\text{(6.12)} \qquad \mathbf{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}') \doteq -\frac{1}{4\pi k^2} \nabla\nabla' \frac{1}{|\mathbf{r} - \mathbf{r}'|} + i\frac{\varepsilon k}{6\pi} + \cdots, \qquad k \to 0$$

that the cross section (7) becomes

$$\text{(6.13)} \qquad \sigma(\mathbf{n}) \doteq \frac{4\pi k^4}{3} \cdot \frac{(\int_{S_1}\{\mathbf{h}\cdot\mathbf{ll}\phi_2(\varrho)+\mathbf{h}\cdot\mathbf{mm}\phi_3(\varrho)\}dS)^2(\int_{S_1}\{(\mathbf{h}\cdot\mathbf{l})^2\phi_2(\varrho)+(\mathbf{h}\cdot\mathbf{m})^2\phi_3(\varrho)\}dS)^2}{\left(\int_{S_1}\{\mathbf{h}\cdot\mathbf{ll}\phi_2(\varrho)+\mathbf{h}\cdot\mathbf{mm}\phi_3(\varrho)\}\cdot\nabla\nabla'\dfrac{1}{|\varrho-\varrho'|}\cdot\{\mathbf{h}\cdot\mathbf{ll}\phi_2(\varrho')+\mathbf{h}\cdot\mathbf{mm}\phi_3(\varrho')\}dS\,dS'\right)^2},$$

$$k \to 0, \qquad \mathbf{e}_z\cdot\mathbf{e} = 0$$

The latter expression reveals a λ^{-4} wave length dependence, characteristic of Rayleigh scattering. A simplification of the rather involved angular dependence occurs, for example, in the case of a circular aperture, where $\phi_2(\varrho) = \phi_3(\varrho) = \phi(\rho)$, and the principal axes coincide with any pair of perpendicular radii; it turns out that

$$\text{(6.14)} \qquad \sigma(\mathbf{n}) \doteq \frac{16\pi k^4}{3} \frac{(\int_{S_1}\phi(\rho)\,dS)^4}{(\int_{S_1}\phi(\rho)(\partial_x^2+\partial_y^2)\,[1/(|\varrho-\varrho'|)]\,\phi(\rho')\,dS\,dS')^2},$$

$$k \to 0, \quad \mathbf{e}_z\cdot\mathbf{e} = 0 \qquad S_1 : \begin{array}{l} 0 \le \rho \le a \\ 0 \le \phi \le 2\pi. \end{array}$$

The cross section (14) may be compared with a corresponding result [8, I, eq. 3.7] in the problem of scalar diffraction theory, where the wave function vanishes at the screen, although the aperture has any shape; we find that for a circular aperture, the electromagnetic cross section is the larger by a factor of 8.

With arbitrary incident plane wave polarization, the cross section has a more complicated angle dependence, although the factor λ^{-4} is retained. It is simpler to obtain the angular features via the diffracted intensity,

$$\text{(6.15)} \qquad \sigma(\mathbf{n}) = \int_0^{\pi/2} d\vartheta' \int_0^{2\pi} d\phi' \sin\vartheta' \, |\mathbf{n}' \times \mathbf{A}(\mathbf{n}', \mathbf{n})|^2,$$

once the amplitude $\mathbf{A}$ is derived from the tangential electric aperture field.

At very short wave lengths, or high frequencies, the problem again simplifies, based on the nearly geometrical nature of wave propagation. This characteristic finds expression in the trial functions for (7),

$$\mathbf{e}_z \times \mathbf{E}_{\pm\mathbf{n}}(\boldsymbol{\varrho}) = \mathbf{e}_z \times \mathbf{e}\Phi_{\mathbf{n}}(\boldsymbol{\varrho}) \exp \{\pm ik\mathbf{n}\cdot\boldsymbol{\varrho}\} \tag{6.16}$$

which differ from the incident fields only by a real modulation factor $\Phi_{\mathbf{n}}(\boldsymbol{\varrho})$, assumed to be frequency independent or of limited variation in distances comparable to the wave length. It follows that

$$\begin{aligned}\sigma(\mathbf{n}) = -\frac{1}{2k}\cos^2\vartheta\left(\int_{S_1}\Phi_{\mathbf{n}}(\boldsymbol{\varrho})\,dS\right)^2\cdot\mathcal{I}m\Big[\int_{S_1}\mathbf{e}_z\times\mathbf{e}\Phi_{\mathbf{n}}(\boldsymbol{\varrho})\exp\{ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\\ \cdot\,\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\cdot\mathbf{e}_z\times\mathbf{e}\Phi_{\mathbf{n}}(\boldsymbol{\varrho}')\exp\{-ik\mathbf{n}\cdot\boldsymbol{\varrho}'\}\,dS\,dS'\Big]^{-1}\end{aligned} \tag{6.17}$$

noting

$$\mathbf{h}\cdot\mathbf{e}_z\times\mathbf{e} = \mathbf{h}\cdot\mathbf{e}_z\times(\mathbf{h}\times\mathbf{n}) = \mathbf{n}\cdot\mathbf{e}_z = \cos\vartheta.$$

A reduction of the multiple integral in (17), with asymptotic validity for $k\to\infty$, is possible on account of the attendant rapid exponential oscillations. This feature suggests extension of the integration domain for the primed coordinates (say) to the entire plane $S_1 + S_2$, and identification of arguments for the $\Phi_{\mathbf{n}}$ functions, the last since oscillations of the integrand are least rapid if $\boldsymbol{\varrho}'$, $\boldsymbol{\varrho}$ are close together. Hence

$$\int_{S_1}\mathbf{e}_z\times\mathbf{e}\Phi_{\mathbf{n}}(\boldsymbol{\varrho})\exp\{ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\cdot\boldsymbol{\Gamma}^{(0)}(\boldsymbol{\varrho},\boldsymbol{\varrho}')\cdot\mathbf{e}_z\times\mathbf{e}\Phi_{\mathbf{n}}(\boldsymbol{\varrho}')\exp\{-ik\mathbf{n}\cdot\boldsymbol{\varrho}'\}\,dS\,dS'$$

$$\simeq\int_{S_1}dS\,\mathbf{e}_z\times\mathbf{e}\Phi_{\mathbf{n}}^2(\boldsymbol{\varrho})\exp\{ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\cdot\left(\boldsymbol{\varepsilon}+\frac{1}{k^2}\nabla\nabla\right)$$

$$\cdot\int_{S_1+S_2}\frac{\exp\{ik\,|\,\boldsymbol{\varrho}-\boldsymbol{\varrho}'\,|\}}{4\pi\,|\,\boldsymbol{\varrho}-\boldsymbol{\varrho}'\,|}\exp\{-ik\mathbf{n}\cdot\boldsymbol{\varrho}'\}\,\mathbf{e}_z\times\mathbf{e}\,dS'$$

$$=\frac{i}{2k\cos\vartheta}\int_{S_1}\Phi_{\mathbf{n}}^2(\boldsymbol{\varrho})\exp\{ik\mathbf{n}\cdot\boldsymbol{\varrho}\}(\mathbf{e}_z\times\mathbf{e})\cdot\left(\boldsymbol{\varepsilon}+\frac{1}{k^2}\nabla\nabla\right)$$

$$\cdot\exp\{-ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\cdot(\mathbf{e}_z\times\mathbf{e})\,dS$$

$$=\frac{i}{2k\cos\vartheta}\int_{S_1}\Phi_{\mathbf{n}}^2(\boldsymbol{\varrho})\mathbf{e}_z\times(\mathbf{h}\times\mathbf{n})\cdot(\boldsymbol{\varepsilon}-\mathbf{nn})\cdot\mathbf{e}_z\times(\mathbf{h}\times\mathbf{n})\,dS$$

$$=\frac{i}{2k}\cos\vartheta\int_{S_1}\Phi_{\mathbf{n}}^2(\boldsymbol{\varrho})\,dS,$$

and thus

$$\sigma(\mathbf{n})\simeq\cos\vartheta\,\frac{(\int_{S_1}\Phi_{\mathbf{n}}(\boldsymbol{\varrho})\,dS)^2}{\int_{S_1}\Phi_{\mathbf{n}}^2(\boldsymbol{\varrho})\,dS},\qquad k\to\infty. \tag{6.18}$$

A stationary value of (18) is attained with Φ_n = const., and the resulting geometrical cross section equals the projected area of the aperture on a plane normal to the direction of the incident wave.

From the Kirchhoff cross section (9), we find

$$\sigma_K(\mathbf{n}) \doteq \frac{(kS_1)^2}{3\pi} (\mathbf{e}_z \times \mathbf{h})^2, \qquad k \to 0 \tag{6.19}$$

the wave length and angle dependence at variance with previous results. However, for short wave lengths, arguments similar to those used in the derivation of (18) yield

$$\sigma_K(\mathbf{n}) \simeq S_1 \cos \vartheta, \qquad k \to \infty \tag{6.20}$$

in accord with the stationary value of (18); the second term of (8) vanishes in this limit, independently of $\mathbf{K}^+$, as revealed by the mutually exclusive integration domains for the numerator. A long wave length approximation for the latter term, adequate to correct the Kirchhoff deficiencies, requires a precise description of $\mathbf{K}^+$, and is therefore complicated.

7. *Diffraction by a Circular Aperture*

To illustrate the variational techniques, we consider the diffraction of plane waves by a circular aperture, with the restriction of normal incidence. By geometrical symmetry, the tangential electric aperture field has a component along the incident electric polarization direction only, and is a function of the radial coordinate. Hence, if $\mathbf{A} \equiv \mathbf{A}(0, 0)$, we learn from (4.11) that

$$A_x = A_y = -\frac{1}{2\pi} \frac{(\int_{S_1} \phi(\rho)\, dS)^2}{\int_{S_1} \phi(\rho)[\Gamma_{xx}^{(0)}(\boldsymbol{\rho}, \boldsymbol{\rho}') + \Gamma_{yy}^{(0)}(\boldsymbol{\rho}, \boldsymbol{\rho}')]\phi(\rho')\, dS\, dS'} . \tag{7.1}$$

An appropriate expansion for $\phi(\rho)$ is

$$\phi(\rho) = \sum_{n=1}^{\infty} A_n \left(1 - \frac{\rho^2}{a^2}\right)^{n-1/2} \tag{7.2}$$

where a is the aperture radius, and the A_n are arbitrary coefficients; the leading term of (2) has a form predicted by the low frequency integral equation. Denoting

$$B_n = \int_{S_1} \left(1 - \frac{\rho^2}{a^2}\right)^{n-1/2} dS = \frac{2\pi a^2}{2n+1}, \tag{7.3}$$

and

$$\begin{aligned} C_{mn} &= \int_{S_1} \left(1 - \frac{\rho^2}{a^2}\right)^{m-1/2} (\Gamma_{xx}^{(0)}(\boldsymbol{\rho}, \boldsymbol{\rho}') + \Gamma_{yy}^{(0)}(\boldsymbol{\rho}, \boldsymbol{\rho}'))\left(1 - \frac{\rho'^2}{a^2}\right)^{n-1/2} dS\, dS' \\ &= C_{nm}, \end{aligned} \tag{7.4}$$

we get

$$A_x \sum_{m,n=1}^{\infty} A_m A_n C_{mn} = -\frac{1}{2\pi}\Big(\sum_{n=1}^{\infty} A_n B_n\Big)^2. \tag{7.5}$$

The result of differentiating with respect to A_m(say), and invoking the stationary property of A_x yields, after some manipulation

$$A_x = -\frac{1}{2\pi}\sum_{n=1}^{\infty} B_n D_n \tag{7.6}$$

where

$$\sum_{n=1}^{\infty} C_{mn} D_n = B_m\,, \qquad m = 1, \cdots . \tag{7.7}$$

If the linear equations (7) are reduced to a finite set with the same number N of unknowns, an approximate value $A_x^{(N)}$ is deduced from (6). It can be shown that

$$A_x^{(N)} = -\frac{1}{2\pi}\left[\frac{B_1^2}{C_{11}} + \sum_{n=2}^{N}\frac{(\bar{\mathfrak{D}}_n)^2}{\mathfrak{D}_{n-1}\mathfrak{D}_n}\right] \tag{7.8}$$

where $\mathfrak{D}_n$ is the determinant $\| C_{mn} \|$ of n rows and columns, and $\bar{\mathfrak{D}}_n$ is obtained from the latter on replacing the last column by $B_1\,, \cdots B_n$.

Employing the integral representation

$$\begin{aligned} &\frac{\exp\{ik \mid \varrho - \varrho' \mid\}}{4\pi \mid \varrho - \varrho' \mid} \\ &= \frac{1}{4\pi}\int_0^{\infty} J_0(\zeta(\rho^2 + \rho'^2 - 2\rho\rho' \cos(\varphi - \varphi'))^{1/2})\,\frac{\zeta\, d\zeta}{(\zeta^2 - k^2)^{1/2}}\,, \end{aligned} \tag{7.9}$$

$$\arg(\zeta^2 - k^2)^{1/2} = 0, \qquad \zeta > k; \qquad = -\pi/2, \qquad \zeta < k$$

and the Bessel function addition theorem,

$$\begin{aligned} &J_0(\zeta(\rho^2 + \rho'^2 - 2\rho\rho' \cos(\phi - \phi'))^{1/2}) \\ &= \sum_{n=0}^{\infty} (2 - \delta_{0n}) J_n(\zeta\rho) J_n(\zeta\rho') \cos n(\phi - \phi'), \qquad \delta_{pq} = \begin{cases} 0, & p \neq q \\ 1, & p = q \end{cases} \end{aligned} \tag{7.10}$$

we find

$$\Gamma_{xx}^{(0)}(\varrho, \varrho') + \Gamma_{yy}^{(0)}(\varrho, \varrho') = \frac{1}{4\pi}\sum_0^{\infty} (2 - \delta_{0n}) \cos n(\phi - \phi') \int_0^{\infty} \zeta\, d\zeta$$

$$\cdot \left[(\zeta^2 - k^2)^{-1/2} - \frac{1}{k^2}(\zeta^2 - k^2)^{1/2}\right] J_n(\zeta\rho) J_n(\zeta\rho'),$$

whence

$$
\begin{aligned}
C_{mn} &= \pi \int_0^\infty \zeta\, d\zeta \left[(\zeta^2 - k^2)^{-1/2} - \frac{1}{k^2} (\zeta^2 - k^2)^{1/2} \right] \\
&\quad \cdot \int_0^a \rho \left(1 - \frac{\rho^2}{a^2}\right)^{m-1/2} J_0(\zeta\rho)\, d\rho \int_0^a \rho' \left(1 - \frac{\rho'^2}{a^2}\right)^{n-1/2} J_0(\zeta\rho')\, d\rho' \\
&= \frac{\pi}{2} \left(\frac{2}{ka}\right)^{m+n} a^3 \Gamma\left(m + \frac{1}{2}\right) \Gamma\left(n + \frac{1}{2}\right) \\
&\quad \cdot \int_0^\infty [(v^2 - 1)^{-1/2} - (v^2 - 1)^{1/2}] v^{-(m+n)} J_{m+1/2}(kav) J_{n+1/2}(kav)\, dv.
\end{aligned} \tag{7.11}
$$

The two integrals of (11) are simply related, for if

$$
\begin{aligned}
F_{mn}(\alpha) &= \int_0^\infty (v^2 - 1)^{1/2} v^{-(m+n)} J_{m+1/2}(\alpha v) J_{n+1/2}(\alpha v)\, dv \\
&= \alpha^{m+n-2} \int_0^\infty (v^2 - \alpha^2)^{1/2} v^{-(m+n)} J_{m+1/2}(v) J_{n+1/2}(v)\, dv,
\end{aligned} \tag{7.12}
$$

then

$$
\begin{aligned}
G_{mn}(\alpha) &= \int_0^\infty (v^2 - 1)^{-1/2} v^{-(m+n)} J_{m+1/2}(\alpha v) J_{n+1/2}(\alpha v)\, dv \\
&= \alpha^{m+n} \int_0^\infty (v^2 - \alpha^2)^{-1/2} v^{-(m+n)} J_{m+1/2}(v) J_{n+1/2}(v)\, dv \\
&= (m + n - 2) F_{mn}(\alpha) - \alpha F'_{mn}(\alpha),
\end{aligned} \tag{7.13}
$$

where the prime signifies differentiation with respect to the argument. Thus,

$$
\begin{aligned}
C_{mn} &= \frac{\pi}{2} \left(\frac{2}{ka}\right)^{m+n} a^3 \Gamma\left(m + \frac{1}{2}\right) \Gamma\left(n + \frac{1}{2}\right) \\
&\quad \cdot [(m + n - 3) F_{mn}(ka) - ka F'_{mn}(ka)].
\end{aligned} \tag{7.14}
$$

The function $F_{mn}(\alpha)$ has been considered elsewhere [8, I, 4.16 et seq.], with its real and imaginary parts given explicitly for $m, n = 1, 2$. Having this information, we return to equation (8) and obtain the transmission coefficient (transmission cross section/area of aperture)

$$
\begin{aligned}
t^{(N)} &= \frac{\sigma^{(N)}}{\pi a^2} = \frac{2}{ka^2} \mathcal{I}m\, A_x^{(N)} \\
&= t^{(1)} - \frac{1}{\pi k a^2} \mathcal{I}m \sum_{n=2}^{N} \frac{(\mathfrak{D}_n)^2}{\mathfrak{D}_{n-1} \mathfrak{D}_n},
\end{aligned} \tag{7.15}
$$

where

$$t^{(1)} = \frac{8}{9\pi} ka \, \mathscr{I}m \frac{1}{F_{11}(ka) + kaF'_{11}(ka)}, \tag{7.16}$$

$$t^{(2)} = \frac{8}{9\pi} ka$$
$$\cdot \mathscr{I}m \left[\frac{F_{22}(ka) - kaF'_{22}(ka) - (1/25)(ka)^2\{F_{11}(ka) + kaF'_{11}(ka) - 10F'_{12}(ka)\}}{\{F_{11}(ka) + kaF'_{11}(ka)\}\{F_{22}(ka) - kaF'_{22}(ka)\} + \{kaF'_{12}(ka)\}^2} \right], \tag{7.17}$$

etc. Curves illustrating the variation of $t^{(1)}$ and $t^{(2)}$ for $0 < ka < 9$ are presented in Figure 2. An expansion in powers of ka gives

$$t^{(1)} = \frac{64}{27\pi^2} (ka)^4 \left[1 + \frac{27}{25} (ka)^2 + 0.72955(ka)^4 + \cdots \right] \tag{7.18}$$

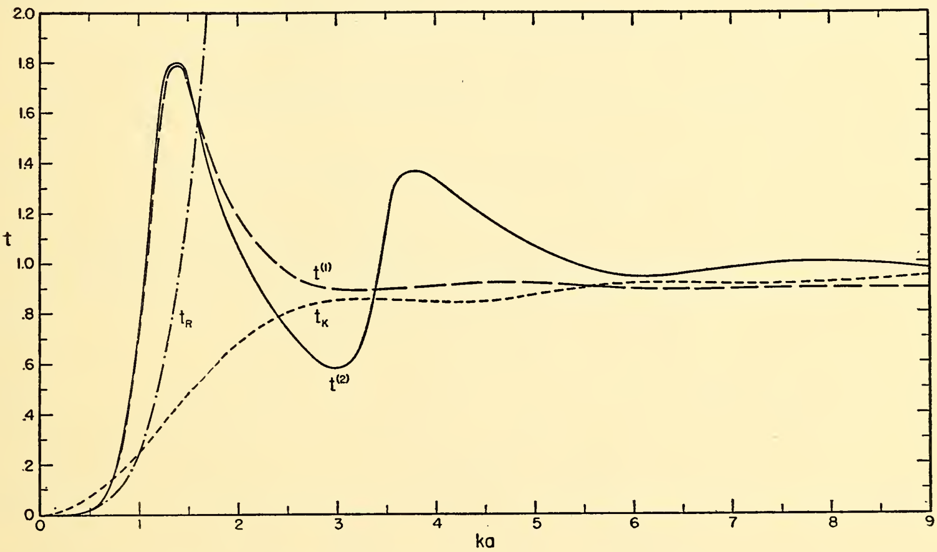


FIG. 2. Transmission coefficient of circular aperture for normal incidence of plane electromagnetic waves. a = radius of aperture, $k = 2\pi/(\text{wave length})$. $t^{(1)}$, first variational approximation, based on tangential electric aperture field of the form $(1 - (\rho^2/a^2))^{1/2}$. $t^{(2)}$, second variational approximation, based on tangential electric aperture field of the form $A_1(1 - (\rho^2/a^2))^{1/2} + A_2(1 - (\rho^2/a^2))^{3/2}$. t_K, Kirchhoff approximation, using Stratton-Chu formulation. t_R, Rayleigh-Bethe approximation.

which confirms the rapid initial rise of transmission coefficient as the parameter increases from very small values; this feature makes it evident that the leading

term of (18), or Rayleigh-Bethe result, is accurate only at extremely long wave lengths. Furthermore,

$$t^{(2)} = \frac{64}{27\pi^2}(ka)^4\left[1 + \frac{27}{25}(ka)^2 + 0.74155(ka)^4 + \cdots\right], \tag{7.19}$$

which differs from (18) in the term of relative order $(ka)^4$. Each approximation $t^{(N)}$ gives correctly the numerical factors for powers of ka less than the square of that by which the corresponding electric aperture field is in error, both relative to the lowest powers; in particular, $t^{(1)}$ is exact through terms of power $(ka)^6$.*

In the short wave length limit, the transmission coefficient is conveniently obtained from (6.18), with the result

$$t^{(N)} \simeq 1 - \frac{1}{(2N+1)^2}, \qquad ka \to \infty \tag{7.20}$$

as for the scalar diffraction problem of [8, I].

It is of interest to compare the variational predictions with those yielded by Kirchhoff approximations. The Kirchhoff amplitude $\mathbf{A}_K(\mathbf{n}, \mathbf{n}')$ which stems from (4.7) on identification of aperture and incident electric fields is

$$\begin{aligned} \mathbf{A}_K(\mathbf{n}, \mathbf{n}') &= \frac{ik}{2\pi}\int_{S_1} \mathbf{e}_z \times \mathbf{E}^{\mathrm{inc}}_{\mathbf{n}'}(\boldsymbol{\varrho}) \exp\{-ik\mathbf{n}\cdot\boldsymbol{\varrho}\}\, dS \\ &= \frac{ik}{2\pi}(\mathbf{e}_z \times \mathbf{e}') \int_{S_1} \exp\{ik(\mathbf{n}' - \mathbf{n})\cdot\boldsymbol{\varrho}\}\, dS; \end{aligned} \tag{7.21}$$

thus, for electric polarization along the x-axis and a circular domain S_1,

$$\mathbf{A}_K(\mathbf{n}, 0) = \mathbf{e}_y\left(2\pi \frac{a}{k} \csc\vartheta J_1(ka \sin\vartheta)\right),$$

so that, using (6.15),

$$\begin{aligned} t_K &= 1 - \frac{1}{2ka}\int_0^{2ka} J_0(t)\, dt \\ &= (ka)^2/3, \qquad ka \to 0; \qquad \simeq 1, \qquad ka \to \infty. \end{aligned} \tag{7.22}$$

Furthermore, the Kirchhoff transmission coefficient obtained from (6.9), based on the infinite screen current distribution, agrees completely with (22). Using

*J. Meixner and W. Andrejewski, Annalen der Physik, Volume 7, 157 (1950) have derived the result

$$t = (64/27\pi^2)(ka)^4[1 + 22/25(ka)^2 + \cdots]$$

from a calculation using spheroidal functions.

the Stratton-Chu formulation, which involves the incident electric and magnetic fields in the aperture and on its rim, it turns out that

$$t_K = 1 - \frac{1}{2}(J_0^2(ka) + J_1^2(ka)) - \frac{1}{4ka}\int_0^{2ka} J_0(t)\,dt \tag{7.23}$$

$$\doteq (7/24)(ka)^2, \qquad ka \to 0; \qquad \simeq 1, \qquad ka \to \infty$$

which differs slightly from (22). According to the numerical results, all Kirchhoff transmission coefficients fail to account for the actual resonance behavior at wave lengths comparable to the aperture dimension, and moreover give a wrong order of magnitude at longer wave lengths. It is therefore evident that the variational formulations enjoy considerable advantage for the practical analysis of diffraction problems.

Appendix I

The Tensor Green's Function of Infinite Empty Space

The free space tensor Green's function is defined as a solution of

$$\nabla \times (\nabla \times \mathbf{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}')) - k^2\mathbf{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}') = \boldsymbol{\varepsilon}\delta(\mathbf{r} - \mathbf{r}') \tag{A.1}$$

upon which are imposed the requirements that all of its components vanish at infinity, and that it satisfy the radiation condition. Its construction is facilitated by first evaluating the divergence of the differential equation it obeys, which yields

$$k^2\nabla\cdot\mathbf{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}') = -\nabla\delta(\mathbf{r} - \mathbf{r}') = \nabla'\delta(\mathbf{r} - \mathbf{r}'), \tag{A.2}$$

and then employing the vector identity

$$\nabla \times (\nabla \times \quad) = \nabla(\nabla\cdot \quad) - \nabla^2(\quad), \tag{A.3}$$

to obtain

$$(\nabla^2 + k^2)\mathbf{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}') = -\left(\boldsymbol{\varepsilon} - \frac{1}{k^2}\nabla\nabla'\right)\delta(\mathbf{r} - \mathbf{r}'). \tag{A.4}$$

The latter equation can be satisfied by writing

$$\mathbf{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}') = \left(\boldsymbol{\varepsilon} - \frac{1}{k^2}\nabla\nabla'\right)G(\mathbf{r}, \mathbf{r}'), \tag{A.5}$$

provided the scalar function G obeys

$$(\nabla^2 + k^2)G(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \tag{A.6}$$

The divergence equation (2) imposes an essential restriction on G, for

$$k^2\nabla\cdot\mathbf{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}') = \nabla'\delta(\mathbf{r} - \mathbf{r}') + k^2(\nabla + \nabla')G(\mathbf{r}, \mathbf{r}'), \tag{A.7}$$

and therefore

$$(\nabla + \nabla')G(\mathbf{r}, \mathbf{r}') = 0, \tag{A.8}$$

which states that $G(\mathbf{r}, \mathbf{r}')$ must be a function of $\mathbf{r} - \mathbf{r}'$; the well known solution of the differential equation for G,

$$G(\mathbf{r}, \mathbf{r}') = \frac{\exp\{ik\,|\,\mathbf{r} - \mathbf{r}'\,|\}}{4\pi\,|\,\mathbf{r} - \mathbf{r}'\,|} \tag{A.9}$$

is indeed a function only of the distance between the two points $\mathbf{r}$ and $\mathbf{r}'$. The choice of sign in the preceding exponential is in deference to the radiation condition, which requires spherical waves moving outward from the source at $\mathbf{r}'$. Hence

$$\mathbf{\Gamma}^{(0)}(\mathbf{r}, \mathbf{r}') = \left(\boldsymbol{\varepsilon} - \frac{1}{k^2}\nabla\nabla'\right)\frac{\exp\{ik\,|\,\mathbf{r} - \mathbf{r}'\,|\}}{4\pi\,|\,\mathbf{r} - \mathbf{r}'\,|} \tag{A.10}$$

is the tensor Green's function of infinite space, for it satisfies the differential equation, the radiation condition, and evidently all of its components vanish at infinity.

Appendix II

Low Frequency Aperture Electric Field

To study the aperture electric field in the low frequency approximation, we start from the integral equation (4.3), written as

$$\begin{aligned} \mathbf{e}_z \times \mathbf{H}_{0n}(\boldsymbol{\varrho}) &= -4ik\int_{S_1} \mathbf{e}_z \times \mathbf{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}')\cdot(\mathbf{e}_z \times \mathbf{E}_n(\boldsymbol{\varrho}'))\,dS' \\ &= -4ik\int_{S_1} \mathbf{e}_z \times \mathbf{\Gamma}^{(0)}(\boldsymbol{\varrho}, \boldsymbol{\varrho}') \times \mathbf{e}_z\cdot\mathbf{E}_{tn}(\boldsymbol{\varrho}')\,dS', \end{aligned} \qquad \boldsymbol{\varrho} \text{ in } S_1 \tag{A.11}$$

where the subscript t denotes tangential (or transverse to z) character. Introducing the dyadic $\mathbf{\Gamma}^{(0)}$ explicitly, the latter equation becomes

$$\begin{aligned}\mathbf{e}_z \times \mathbf{H}_{0n}(\varrho) &= \frac{ik}{\pi}\int_{S_1} \mathbf{E}_{tn}(\varrho')\,\frac{\exp\{ik\,|\,\varrho-\varrho'\,|\}}{|\,\varrho-\varrho'\,|}\,dS' \\ &\quad - \frac{i}{\pi k}\int_{S_1} \mathbf{e}_z \times \nabla_t\nabla_t\,\frac{\exp\{ik\,|\,\varrho-\varrho'\,|\}}{|\,\varrho-\varrho'\,|} \times \mathbf{e}_z\cdot\mathbf{E}_{tn}(\varrho')\,dS',\end{aligned} \tag{A.12}$$

and by use of the vector identity

$$\mathbf{e}_z \times \nabla_t\nabla_t \times \mathbf{e}_z = \nabla_t \times (\nabla_t \times \quad) = \nabla_t\nabla_t - \boldsymbol{\varepsilon}_t\nabla_t^2\,, \tag{A.13}$$

it follows that

$$\begin{aligned}(\nabla_t \times \nabla_t \times -k^2)\int_{S_1} \mathbf{E}_{tn}(\varrho')\,\frac{\exp\{ik\,|\,\varrho-\varrho'\,|\}}{|\,\varrho-\varrho'\,|}\,dS' & \\ = i\pi k\mathbf{e}_z \times \mathbf{H}_{0n}(\varrho) & \\ = 2i\pi k\mathbf{e}_z \times \mathbf{h}\exp\{ik\mathbf{n}\cdot\varrho\} & \qquad \varrho \text{ in } S_1\,.\end{aligned} \tag{A.14}$$

At low frequencies, the aperture electric field has a twofold composition, of magnetic ($\mathbf{H}_0$) and electric ($\mathbf{E}_0$) origin, respectively. The magnetic part is determined by the integro-differential equation (14) on setting $k = 0$ everywhere in its left hand side and in $\mathbf{H}_{0n}$ on the right hand side; the electric part is given by a corresponding alteration of the equation which results on taking the divergence of (14), namely

$$\begin{aligned}\nabla_t\cdot\int_{S_1} \mathbf{E}_{tn}(\varrho')\,\frac{\exp\{ik\,|\,\varrho-\varrho'\,|\}}{|\,\varrho-\varrho'\,|}\,dS' & \\ = \frac{i\pi}{k}\mathbf{e}_z\cdot\nabla \times \mathbf{H}_{0n}(\varrho) & \\ = \pi\mathbf{e}_z\cdot\mathbf{E}_{0n}(\varrho) = 2\pi\mathbf{e}_z\cdot\mathbf{e}\exp\{ik\mathbf{n}\cdot\varrho\}. &\end{aligned} \tag{A.15}$$

Consequently, the basic low frequency equations are

$$\nabla_t \times \left(\nabla_t \times \int_{S_1} \mathbf{E}_t(\varrho')\,\frac{dS'}{|\,\varrho-\varrho'\,|}\right) = 2\pi ik\mathbf{e}_z \times \mathbf{h} \tag{A.16}$$

and $\qquad k \to 0, \quad \varrho \text{ in } S_1$

$$\nabla_t\cdot\int_{S_1} \mathbf{E}_t(\varrho')\,\frac{dS'}{|\,\varrho-\varrho'\,|} = 2\pi\mathbf{e}_z\cdot\mathbf{e} \tag{A.17}$$

omitting a subscript which refers to the incident plane wave propagation direction;

A further investigation reveals that the entire aperture field has a similar

decomposition, with a magnetic part in which the normal magnetic field predominates, and an electric part with predominant tangential electric field. For each type of excitation, there is an associated dipole moment, magnetic in the plane of the aperture and electric normal to the latter. According to the usual formulas, the magnetic dipole moment is proportional to the integral of $\mathbf{e}_z \times \mathbf{E}(\boldsymbol{\varrho})$, or magnetic current density, over the aperture, while for the electric dipole moment the integral of $(\mathbf{e}_z \times \mathbf{E}(\boldsymbol{\varrho})) \times \boldsymbol{\varrho}$ is involved.

As regards the form of the aperture field, we note that its tangential electric component may be written

$$\text{(A.18)} \qquad \mathbf{E}_E(\boldsymbol{\varrho}) = \mathbf{e}_z \cdot \mathbf{e} \nabla \phi_1(\boldsymbol{\varrho})$$

where ϕ_1 is a real, frequency independent function; it is readily verified that (18) makes no contribution to (16), provided ϕ_1 vanishes on the aperture rim. For the tangential magnetic component, we have

$$\text{(A.19)} \qquad \mathbf{E}_H(\boldsymbol{\varrho}) = ik(-\mathbf{h}\cdot\mathbf{l}\mathbf{m}\phi_2(\boldsymbol{\varrho}) + \mathbf{h}\cdot\mathbf{m}\mathbf{l}\phi_3(\boldsymbol{\varrho}))$$

where $\mathbf{l}$, $\mathbf{m}$ are unit vectors along the principal axes of the dyadic which relates the magnetic dipole moment to $\mathbf{H}_0$, and

$$\text{(A.20)} \qquad \mathbf{e}_z \times \mathbf{l} = \mathbf{m}, \qquad \mathbf{e}_z \times \mathbf{m} = -\mathbf{l}, \qquad \mathbf{l}\cdot\mathbf{m} = 0;$$

the functions ϕ_2, ϕ_3 are real and frequency independent. To justify (19), observe that the related magnetic moment has components along the principal axes which are proportional to those of $\mathbf{H}_0$. The total field $\mathbf{E}_t(\boldsymbol{\varrho}) = \mathbf{E}_E(\boldsymbol{\varrho}) + \mathbf{E}_H(\boldsymbol{\varrho})$ thus has the form indicated by (6.11).

In particular, for a circular aperture, $\mathbf{l} = \mathbf{e}_x$, $\mathbf{m} = \mathbf{e}_y$ and

$$\phi_1 = -\frac{1}{2}\phi_2 = -\frac{1}{2}\phi_3 = -\frac{2}{\pi}(a^2 - \rho^2)^{1/2}.$$

Addition in Proofs:

The variational results for diffraction by a circular aperture obtained in section 7 require important qualification. This is a consequence of recent investigations by Bouwkamp (to appear in Philips Research Reports), which show that the low frequency aperture electric field for normal incidence has components both parallel and perpendicular to the incident electric polarization direction, and exhibits angular asymmetry. Specifically, if the incident electric polarization is along the x direction, the aperture field components according to Bouwkamp are (omitting scale factors)

$$\text{(A)} \qquad E_x = \frac{2a^2 - x^2 - 2y^2}{(a^2 - x^2 - y^2)^{1/2}}, \qquad E_y = \frac{xy}{(a^2 - x^2 - y^2)^{1/2}}, \qquad ka \to 0$$

or in polar coordinates,

$$\text{(B)} \qquad E_\rho = \frac{(2a^2 - \rho^2)\cos\varphi}{(a^2 - \rho^2)^{1/2}}, \qquad E_\varphi = -2(a^2 - \rho^2)^{1/2} \sin\varphi.$$

With this field, the variational transmission coefficient turns out to be

$$\text{(C)} \qquad t = \frac{16}{\pi} \frac{P(ka)}{P^2(ka) + Q^2(ka)},$$

where

$$P(\alpha) = \frac{9}{\pi} - \frac{9}{2\pi\alpha^2} + \frac{9}{4\alpha^3} S_0(2\alpha) + \frac{9}{2\alpha^2} S_1(2\alpha) + \left(1 - \frac{9}{2\alpha^2} - \frac{9}{8\alpha^4}\right) \int_0^{2\alpha} S_0(t)\, dt,$$

$$Q(\alpha) = \frac{9}{4\alpha^3} J_0(2\alpha) + \frac{9}{2\alpha^2} J_1(2\alpha) + \left(1 - \frac{9}{2\alpha^2} - \frac{9}{8\alpha^4}\right) \int_0^{2\alpha} J_0(t)\, dt,$$

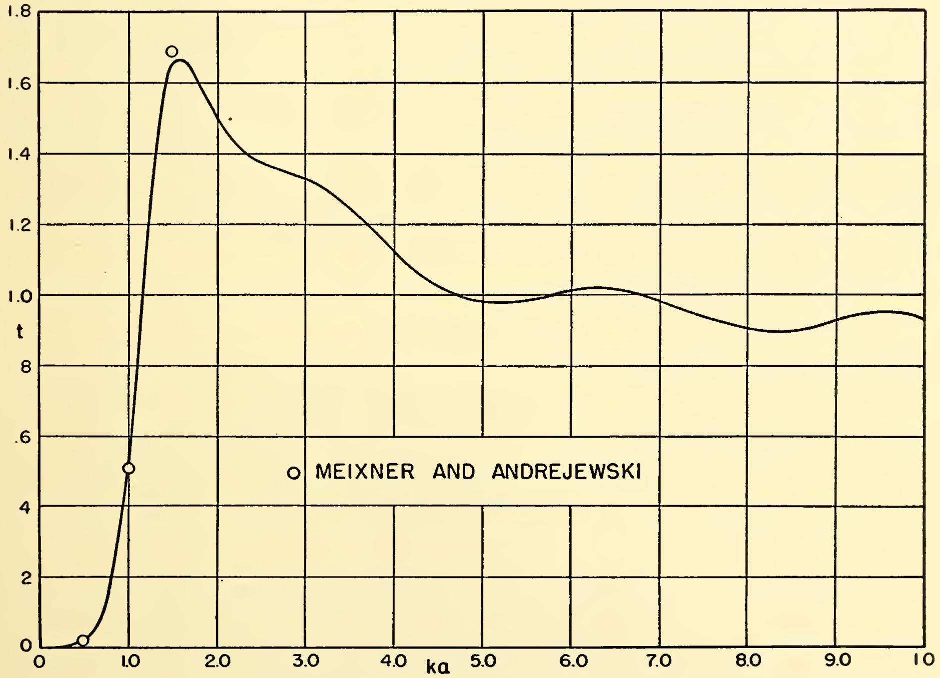


FIG. 3. Variational approximation to the transmission coefficient of a circular aperture, based on Bouwkamp's form of electric aperture field. Points computed from rigorous theory using spheroidal functions.

and J_0, J_1 denote the Bessel functions of order zero and one, while S_0, S_1 denote the corresponding Struve functions.

An expansion of t for small values of ka yields

$$t = \frac{64}{27\pi^2}(ka)^4\left[1 + \frac{22}{25}(ka)^2 + 0.4079(ka)^4 + \cdots\right]$$

which may be compared with the exact result of Bouwkamp (see also Meixner and Andrejewski),

$$t^{\text{exact}} = \frac{64}{27\pi^2}(ka)^4\left[1 + \frac{22}{25}(ka)^2 + 0.3979(ka)^4 + \cdots\right].$$

From agreement of the latter expansions through terms of relative order $(ka)^2$, the correctness of the low frequency aperture field (A) is confirmed. The remarks concerning accuracy of the variational approximations (7.16) and (7.17) in the text are therefore erroneous.

At very high frequencies, $ka \gg 1$, the transmission coefficient in (C) approaches zero, since $P(ka)$ increases logarithmically while $Q(ka)$ tends to unity. A null transmission coefficient can also be inferred from (6.18), for integrals of E_x^2 and E_y^2 over the aperture are infinite, in consequence of the field singularities at the rim of the screen. The latter feature of the low frequency field renders it a poor variational trial function when the frequency becomes very great, as the correct field is then more nearly constant over the aperture (for normal incidence); hence the corresponding variational predictions are inaccurate. From data of the accompanying figure, it can be inferred that the transmission coefficient (C) decreases slowly when $ka \gg 1$, where the alternative formulation based on screen current becomes appropriate.

Bibliography

1. J. A. Stratton and L. J. Chu, *Diffraction theory of electromagnetic waves*, The Physical Review, Volume 56, 1939, p. 99.
2. Lord Rayleigh, *On the incidence of aerial and electric waves on small obstacles in the form of ellipsoids or elliptic cylinders, on the passage of electric waves through a circular aperture in a conducting screen*, The Philosophical Magazine, Volume 44, 1897, p. 28.
3. H. Bateman, *The Mathematical Analysis of Electrical and Optical Wave Motion*, Cambridge, 1915, p. 90.
4. H. A. Bethe, *Theory of diffraction by small holes*, The Physical Review, Volume 66, 1944, p. 163. Also E. T. Copson, *An integral equation method of solving plain diffraction problem*, Proceedings of the Royal Society (Series A), Volume 186, 1946, p. 100; and W. R. Smythe, *The double current sheet in diffraction*, The Physical Review, Volume 72, 1947, p. 1066.
5. Joseph Meixner, *Strenge Theorie der Beugung elektromagnetischer Wellen an der vollkommen leitenden Kreisscheibe*, Zeitschrift für Naturforschung, Volume 3A, 1948, p. 506.
6. P. Debye, *Der Lichtdruck auf Kugeln von Beliebigen Material*, Annalen der Physik (Series 4), Volume 30, 1909, p. 57.

7. C. J. Bouwkamp, *A note on singularities occurring at sharp edges in electromagnetic diffraction theory*, Physica, Volume 12, 1946, p. 467.
8. H. Levine and J. Schwinger, *On the theory of diffraction by an aperture in an infinite plane screen*, Part I, Physical Review, Volume 74, 1948, p. 958. Part II, The Physical Review, Volume 75, 1949, p. 1423.
9. J. W. Miles, *On the diffraction of an electromagnetic wave through a plane screen*, Journal of Applied Physics, Volume 20, 1949, p. 760. *Errata*, Journal of Applied Physics, Volume 21, 1950, p. 468. Miles has also considered variational aspects of the electromagnetic diffraction problem. The power transmitted through the aperture and a related aperture impedance are presented in stationary forms; however, stationary properties of the diffracted field amplitudes are not discussed.

On Systems of Linear Equations in the Theory of Guided Waves

By W. MAGNUS and F. OBERHETTINGER
California Institute of Technology

Introduction and Summary

We investigate the diffraction of an electromagnetic wave between two parallel planes or in a wave guide of rectangular cross section by a plane strip. We assume that all components of the field depend on the time t in a purely periodic way which can be described by a factor exp $\{i\omega t\}$, and that the frequency ω and the proportions of the wave guide are such that there exists essentially only one type of waves which is not attenuated. If the incoming wave is of the type exp $\{i\alpha x\}$ cos βy where α, β are real, the components of the diffracted wave can be expanded in a Fourier series. It becomes evident that the coefficients of the series are uniquely determined by the condition of the finiteness of the total energy in any finite part of the space. This condition has already been used by Bouwkamp [1], Maue [6] and Meixner [7] who also showed that the components of the electromagnetic field behave at the edge of a plane diffracting obstacle like $\rho^{n/2}$, where $n = -1, 0, 1, \cdots$ and where ρ is the distance from the edge. The Fourier coefficients are determined by an infinite system of linear equations; for a certain closely related system the existence of a solution was proved recently by Schaefke [11]. If the width of the diffracting strip is exactly one half of the width of the wave guide, the system of linear equations can be dealt with by a process of successive approximation, such that the first steps can be carried through explicitly. The results of Meixner [7] are used as a guiding principle for the successive approximations. The method is discussed in section 5. The mathematical aids are given in the appendix.

Notations

The system of units is the "practical system"; i.e., the electric field strength is measured in volt/cm etc. The frequency ω of the incoming waves defines

---

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Sciences and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directories of the Air Force Cambridge Research Laboratories.

the wave length $\lambda = 2\pi c/\omega$, (c = velocity of light); a, b are the measures of the wave guides, and we define:

$k = k_0 = 2\pi/\lambda$

$k_m = k_0[1 - m^2\lambda^2 b^{-2}]^{1/2}, \qquad m = 0, 1, 2, \cdots$

k_m is negative imaginary if $m \geqq 1$.

$\mu = \lambda/b$ in Part one; $\qquad (\mu > 1)$

$k_{m+1/2} = k[1 - (m + \frac{1}{2})^2\lambda^2 a^{-2}]^{1/2}, \qquad m = 0, 1, 2, \cdots$

$k_{1/2}$ is real and positive, $k_{m+1/2}$ is negative imaginary if

$m \geqq 1$

$\mu = \lambda/a$ in Part two; $\qquad (2 > \mu > 2/3)$

β denotes the relative width of the diffracting strip (as compared with b or a in §2 or §3).

$\delta_{nm} = 0$ if $n \neq m$, $\delta_{n,n} = 1; \qquad n, m = 0, 1, 2, \cdots .$

$I = (\delta_{n,m})$ is the unit-matrix.

$\epsilon_n = 2$ if $n = 1, 2, 3, \cdots ; \qquad \epsilon_0 = 1; \qquad (a)_n = \Gamma(a + n)/\Gamma(a)$

I. A Diffracting Strip between Two Parallel Planes

1. *Elementary Results*

We consider two parallel planes which, in Cartesian coordinates x, y, z are defined by the equations $z = \pm\frac{1}{2}b$. Between the two planes we have an infinite strip defined by $x = 0$, $-\frac{1}{2}\beta b \leqq z \leqq \frac{1}{2}\beta b$, $-\infty < y < \infty$; $0 < \beta < 1$. A cross section of this arrangement is shown in Figure 1.

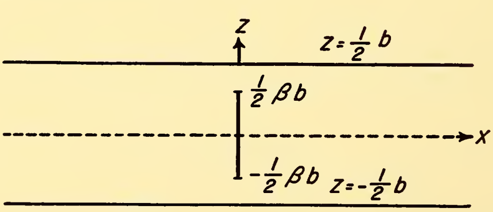


FIGURE 1

We assume that the planes and the strip have infinite electric conductivity and we consider an incoming electromagnetic wave which is defined by

$$E = E_0 e^{-ikx}, \tag{1.1}$$

where E denotes the z-component of the electric field, E_0 is a constant, and where the other components of the electric field vanish. The time factor exp $\{i\omega t\}$ will be omitted everywhere.

We may assume that the x and y components of the diffracted wave vanish and that its z-component E_z can be expanded in a series

$$(1.2)\qquad E_z(x, z) = -60\pi \sum_{m=0}^{\infty} A_m(k_m/k) \cos(2\pi mz/b) \exp\{-i \mid x \mid k_m\}$$

which is convergent for all $x \neq 0$. If the series

$$(1.3)\qquad I(z) = \sum_{m=0}^{\infty} A_m \cos 2\pi mz/b$$

converges, we may call $I(z)$ the current on the strip. We could derive the expression for E_z from $I(z)$ and from the formula for the field of an electric dipole between two parallel planes (cf. for instance [8]).

From the condition that the total energy of the field contained in any finite part of the space should be finite we find that

$$(1.4)\qquad \iint \mid E_z \mid^2 dx\, dz = \mid A_0 \mid^2 + \frac{1}{2} \sum_{m=1}^{\infty} \mid A_m^2 \mid \int_{\epsilon}^{1} \exp\{-2 \mid xk_m \mid\} \mid k_m \mid^2 dx$$

is bounded if $\epsilon \to 0$. Therefore we have

$$(1.5)\qquad \sum_{m=1}^{\infty} \mid A_m^2 \mid \mid k_m \mid < \infty .$$

The boundary conditions are

$$\text{(i)}\qquad E_z(0, z) = -E_0 \qquad \text{for} \qquad -\tfrac{1}{2}\beta b < z < \tfrac{1}{2}\beta b$$

$$\text{(ii)}\qquad I(z) = 0 \qquad \text{for} \qquad \tfrac{1}{2}\beta b < \mid z \mid < \tfrac{1}{2}b;$$

Condition (ii) can be derived from the fact that $I(z)$ is (apart from a constant factor) the y-component of the magnetic field at $x = 0$. It must be regular outside the strip. From Maxwell's equations it follows that $I(z)$ is an odd function of x since the electric field is an even function of x. If $x \to \infty$ we assume: (iii) There exists a constant C such that

$$(1.6)\qquad \lim_{|x| \to \infty} \mid E_z(x, z) - C \exp\{-ik \mid x \mid\} \mid = 0.$$

In the present very simple case this happens to be equivalent to Sommerfeld's "radiation condition [4, 10, 13], because we assume that only $k_0 = k$ is real and that k_m is negative-imaginary for $m = 1, 2, 3, \cdots$.

From (i) and (ii) we find

$$(1.7)\qquad \int_{-\frac{1}{2}b}^{\frac{1}{2}b} I(z)\{\bar{E}_z(0, z) + \bar{E}_0\}\, dz = 0.$$

Since $E_z(x, z)$ must approach $E_z(0, z)$ if $x \to 0$ we deduce from (1.2) and (1.5) that we can introduce the formal series for $\overline{E}_z(0, z)$ and $I(z)$ into (1.7). This gives

$$60\pi A_0 \overline{A}_0 = \Re e A_0 \overline{E}_0 \tag{1.8}$$

$$30\pi \sum_{m=1}^{\infty} |\, k_m \,|\, A_m \overline{A}_m = \Im m\, k A_0 \overline{E}_0 \,. \tag{1.9}$$

Therefore the A_m are uniquely determined by (1.5) and by the boundary conditions. If this were not true, a non-trivial solution of the problem would exist for which $E_0 = 0$; this contradicts (1.8), (1.9). Finally, the complete solution of the problem is uniquely determined by (iii) if the A_m are given.

If $|\, x \,| \to \infty$, only the first term in the expansion (1.2) contributes to E_z. Therefore we may call

$$R = -60\pi A_0/E_0 \tag{1.10}$$

the "reflection coefficient". It can be shown (in an elementary way) that $|\, R \,|$ is always different from 0 and 1 if β (the relative width of the diffracting strip) is different from 0 and 1. But the inequalities for $|\, R \,|$ which can be obtained by an elementary method are unsatisfactory.

2. *The Linear Equations for* $\beta = \frac{1}{2}$

From (1.5) it does not follow that the expansion (1.2) is convergent if $x = 0$. In order to obtain a system of linear equations for the A_m we shall assume that this is the case. This assumption can be justified to some extent by the results in [7]. According to Meixner,

$$\lim E_z(0, z)[\tfrac{1}{4}\beta^2 b^2 - z^2]^{1/2}$$

exists if $z = \pm\frac{1}{2}\beta b \pm \epsilon$ and $\epsilon \to 0$. Since we can construct a convergent Fourier series for a function which is constant in one of the domains $\frac{1}{2}\beta b < |\, z \,| < \frac{1}{2} b$ and $-\frac{1}{2}\beta b < z < \frac{1}{2}\beta b$ and is equal to

$$[\tfrac{1}{4}\beta^2 b^2 - z^2]^{-1/2}$$

in the other domain, we may expect that $E_z(0, z)$ can be expanded in a series of this type plus the Fourier series of an L^2-function which is at least of the class C'' except at $z = \pm\frac{1}{2}\beta b$. This and the formulas (A.18), (A.19) of the appendix lead to the assumption that

$$|\, A_m m^{3/2} \,| \leqq M \tag{2.1}$$

where M is independent of m.

Normalization. We denote z/b by ζ and we take $E_0 = 60\pi$. Then $-A_0$ is the reflection coefficient. We define a complete orthonormal system of even functions $\phi_m(\zeta)$ in $(-\frac{1}{2}, \frac{1}{2})$ by

$$\phi_0(\zeta) = 1, \qquad \phi_m(\zeta) = \sqrt{2} \cos 2\pi m\zeta, \qquad m = 1, 2, 3, \cdots \tag{2.2}$$

According to H. L. Schmid's lemma (cf. Appendix I) we find that the A_n have to be computed from

$$A_n = \sum_{m=0}^{\infty} u_{n,m} x_m; \qquad u_{n,m} = \int_{-\frac{1}{2}\beta}^{\frac{1}{2}\beta} \phi_n(\zeta)\phi_m(\zeta)\, d\zeta, \tag{2.3}$$

where the x_m are determined by the equations

$$\frac{k_n}{k} \sum_{m=0}^{\infty} u_{n,m} x_m + \sum_{m=0}^{\infty} (\delta_{n,m} - u_{n,m}) x_m = \delta_{0,n}, \qquad n = 0, 1, 2, \cdots . \tag{2.4}$$

Restriction to $\beta = \frac{1}{2}$, and Approximation. If $\beta = \frac{1}{2}$, we may write $u_{n,m} = \frac{1}{2}\delta_{n,m} + \frac{1}{2}s_{n,m}$, where the matrix $S = (s_{n,m})$ satisfies $S^2 = I$. Introducing

$$\gamma_l = \sqrt{\epsilon_l}\,(-1)^l x_{2l}, \qquad \eta_r = (-1)^r x_{2r+1}, \qquad l, r = 0, 1, 2, \cdots \tag{2.5}$$

and computing the $s_{n,m}$ from (2.3) we find that (2.4) can be written in the form

$$\gamma_0 = 1 \tag{2.6}$$

$$\tau_{2l}\gamma_l + \frac{1}{\pi\sqrt{2}} \sum_{r=0}^{\infty} \left(\frac{2}{l + r + \frac{1}{2}} + \frac{2}{-l + r + \frac{1}{2}} \right) \eta_r = 0, \quad l = 1, 2, 3, \cdots \tag{2.7}$$

$$\tau_{2r+1}\eta_r + \frac{1}{\pi\sqrt{2}} \sum_{l=0}^{\infty} \left(\frac{1}{l + r + \frac{1}{2}} + \frac{1}{-l + r + \frac{1}{2}} \right) \gamma_l = 0, \quad r = 0, 1, \cdots \tag{2.8}$$

where

$$\tau_m = \frac{k_m + k}{k_m - k} = 1 + \frac{2i}{m\mu}\,[1 - m^{-2}\mu^{-2}]^{1/2} - \frac{2}{m^2\mu^2}; \qquad \left(\mu = \frac{\lambda}{b}\right), \qquad m = 1, 2, 3, \cdots . \tag{2.9}$$

We can express the γ_l in terms of the η_r by applying Titchmarsh's inversion formula (cf. Appendix I, (A.7), (A.8)) to (2.8). This gives

$$\gamma_l = -\frac{\epsilon_l}{\pi\sqrt{2}} \sum_{r=0}^{\infty} \left(\frac{1}{-l + r + \frac{1}{2}} + \frac{1}{l + r + \frac{1}{2}} \right) \tau_{2r+1}\eta_r \tag{2.10}$$

$$l = 0, 1, 2, \cdots; \qquad \epsilon_0 = 1; \qquad \epsilon_1 = \epsilon_2 = \cdots = 2.$$

Eliminating the γ_l, we find from (2.6), (2.9), (2.10):

$$\frac{\sqrt{2}}{\pi}\sum_{r=0}^{\infty}\frac{\tau_{2r+1}\eta_r}{r+\frac{1}{2}} = -1 \tag{2.11}$$

$$\sum_{r=0}^{\infty}\left\{\frac{1}{-l+r+\frac{1}{2}}+\frac{1}{l+r+\frac{1}{2}}\right\}(\tau_{2l}^{-1}-\tau_{2r+1})\eta_r = 0, \quad l = 1, 2, 3, \cdots \tag{2.12}$$

Expanding the τ_m in a series of powers of $(m\mu)^{-1}$ and neglecting the terms of a degree greater than two, we obtain the *Approximative set of linear equations*:

$$\begin{cases} \dfrac{1}{\pi}\displaystyle\sum_{r=0}^{\infty}\frac{\eta_r}{r+\frac{1}{2}}\left\{1+\frac{i}{\mu(r+\frac{1}{2})}-\frac{1}{2}\frac{1}{\mu^2(r+\frac{1}{2})^2}\right\} = -\dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\pi}\displaystyle\sum_{r=0}^{\infty}\frac{\eta_r}{-l+r+\frac{1}{2}}\left\{1-\frac{i}{2\mu}\left[\frac{1}{l}-\frac{1}{r+\frac{1}{2}}\right]\right\} = 0, \qquad l = 1, 2, 3, \cdots \end{cases} \tag{2.13}$$

We apply the Linfoot-Shepherd inversion formula for $\theta = \frac{1}{2}$(cf. Appendix I) by multiplying the l-th equation by

$$\left(n+\frac{1}{2}\right)\frac{(\frac{1}{2})_n}{n!}\frac{1}{-l+n+\frac{1}{2}}\frac{(\frac{1}{2})_l}{l!}, \qquad n = 0, 1, 2, \cdots$$

and adding all the equations. This and an application of the formulas in Appendix IV gives

$$\begin{aligned} \eta_n + \frac{1}{\pi}\frac{(\frac{1}{2})_n}{n!}\frac{i}{\mu}&\left\{\sum_{r=0}^{\infty}\eta_r\left[\frac{-\log 2}{r+\frac{1}{2}}+\frac{1}{(r+\frac{1}{2})^2}+\frac{i}{2\mu}\frac{1}{(r+\frac{1}{2})^3}\right]\right. \\ &\left.+\frac{1}{2}\frac{1}{n+\frac{1}{2}}\sum_{r=0}^{\infty}\frac{\eta_r}{r+\frac{1}{2}}\right\} = -\frac{1}{\sqrt{2}}\frac{(\frac{1}{2})_n}{n!}. \end{aligned} \tag{2.14}$$

We can obtain a solution of (2.14) by letting

$$\eta_n = -\frac{1}{\sqrt{2}}\frac{(\frac{1}{2})_n}{n!}\left\{\sigma+\frac{1}{2}\frac{1}{n+\frac{1}{2}}\tau\right\}. \tag{2.15}$$

Defining S and T by

$$S = \sum_{r=0}^{\infty}\eta_r\left\{\frac{-\log 2}{r+\frac{1}{2}}+\frac{1}{(r+\frac{1}{2})^2}+\frac{i}{2\mu}\frac{1}{(r+\frac{1}{2})^3}\right\} \tag{2.16}$$

$$T = \sum_{r=0}^{\infty}\frac{\eta_r}{r+\frac{1}{2}}, \tag{2.17}$$

they become linear functions of σ, τ if we substitute the right side of (2.15) in (2.16), (2.17); then equation (2.14) can be written in the form

$$\sigma+\frac{1}{2}\frac{1}{n+\frac{1}{2}}\tau-\frac{i}{\mu}\frac{\sqrt{2}}{\pi}\left(S+\frac{1}{2}\frac{1}{n+\frac{1}{2}}T\right) = 1. \tag{2.18}$$

This is satisfied (for all values of n) if and only if

$$\sigma - \frac{i}{\mu}\frac{\sqrt{2}}{\pi} S = 1, \qquad \tau - \frac{i}{\mu}\frac{\sqrt{2}}{\pi} T = 0. \tag{2.19}$$

Therefore we obtain two linear equations for σ, τ, which can be written in the form

$$\begin{aligned} &\sigma\left\{1 + \frac{i}{\pi\mu}\left[-(\log 2)S_1 + S_2 + \frac{i}{2\mu} S_3\right]\right\} \\ &\qquad + \frac{i}{\pi\mu}\,\tau\left\{-\frac{1}{2}(\log 2)S_2 + S_3 + \frac{i}{2\mu} S_4\right\} = 1 \end{aligned} \tag{2.20}$$

$$\sigma \frac{i}{\pi\mu} S_1 + \tau\left(1 + \frac{1}{2\pi\mu} S_2\right) = 0 \tag{2.21}$$

where (cf. Appendix IV, A.23)

$$S_n = \sum_{r=0}^{\infty} \frac{(\frac{1}{2})_r}{r!}\left(r + \frac{1}{2}\right)^{-n}, \qquad n = 1, 2, 3, \cdots \tag{2.22}$$

$$S_1 = \pi; \qquad S_2 = 2\pi \log 2; \qquad S_3 = 2\pi\left[(\log 2)^2 - \frac{1}{12}\pi^2\right]. \tag{2.23}$$

Equations (2.20) and (2.21) determine σ, τ uniquely; therefore the η_r are given by (2.15). From these approximate values of the η_r we obtain approximate values of the γ_l from (2.6), (2.7). From (2.3) and from the equations (2.7), (2.8) we can now compute the A_n, $n = 0, 1, 2, \cdots$. This gives

Theorem 1. *If we simplify the linear equations* (2.6), (2.7), (2.8) *by substituting for* τ_m *in* (2.9) *the first three terms of its expansion in a series of powers of* $(\mu m)^{-1}$, *then the Fourier-coefficients of the "current" on the diffracting strip become*

$$A_0 = \tfrac{1}{2} - \tfrac{1}{2}(\sigma + \tau \log 2) \tag{2.24}$$

$$A_{2l} = \frac{(-1)^{l+1}}{\sqrt{2}}\,[1 - i(4l^2\mu^2 - 1)^{1/2}]^{-1}\,\frac{(\frac{1}{2})_l}{l!}\left(\sigma - \frac{1}{2l}\tau\right), \qquad l = 1, 2, 3, \cdots \tag{2.25}$$

$$A_{2r+1} = \frac{(-1)^{r+1}}{\sqrt{2}}\,\{1 + i[(2r+1)^2\mu^2 - 1]^{1/2}\}^{-1}\,\frac{(\frac{1}{2})_r}{r!}\left(\sigma + \frac{1}{2r+1}\tau\right), \qquad r = 0, 1, 2, \cdots \tag{2.26}$$

where σ, τ *are defined by* (2.20), (2.21). For large values of μ we have approximately

$$\sigma = \frac{1}{1 + \mu^{-1} i \log 2} + \mathcal{O}(\mu^{-2}); \qquad \tau = -\frac{i}{\mu}\left(1 + \frac{i}{\mu}\log 2\right)^{-2} + \mathcal{O}(\mu^{-2}).$$

The approximate solutions A_n in equations (2.24), (2.25), (2.26) satisfy (2.1); the reflection coefficient $A_0 \to 0$ if $\mu \to \infty$.

II. A Diffracting Strip in a Wave Guide with a Rectangular Cross Section

3. *Elementary Results*

Let us consider a rectangular wave guide which extends in the direction of the x-axis. The corners of the cross section of the wave guide in the y,z-plane are given by $y = \pm\frac{1}{2}a$, $z = \pm\frac{1}{2}a'$, a diffracting strip occupies the area $-\frac{1}{2}\beta a \leqq y \leqq \frac{1}{2}\beta a$, $-\frac{1}{2}a' \leqq z \leqq \frac{1}{2}a'$, where $0 < \beta < 1$; in §4, we shall choose $\beta = \frac{1}{2}$. Figure 2 shows the cross section of the wave guide.

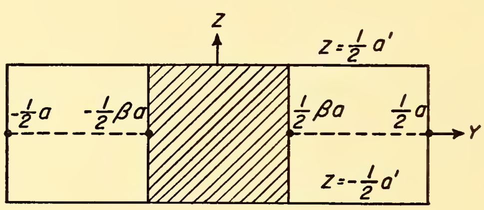


FIGURE 2

We assume that the wave guide and the (infinitely thin) strip have infinite electrical conductivity and we consider an incoming wave which is defined by

$$E = E_0 \cos(\pi y/a) \exp\{-ixk_{1/2}\}, \tag{3.1}$$

where E denotes the z-component of the electric field, E_0 is a constant, and all the other components of the incoming electric field vanish. The time factor $\exp\{i\omega t\}$ will be omitted everywhere.

As in §1, we can show that the diffracted wave has an electric field parallel to the z-axis, and that its z-component can be expanded in a series

$$E_z(x, y) = -60\pi k \sum_{m=0}^{\infty} \frac{A_m}{k_{m+1/2}} \cos(2m + 1)\frac{\pi y}{a} \exp\{-i \mid x \mid k_{m+1/2}\} \tag{3.2}$$

which converges for all $\mid x \mid > 0$ (and even for $x = 0$, as we shall see). If the series

$$I(y) = \sum_{m=0}^{\infty} A_m \cos((2m+1)\pi y/a) \tag{3.3}$$

converges, we can call $I(y)$ the current on the strip and derive the diffracted wave from (3.3) and the formulas for the electric field of a dipole in a wave guide. (cf. [8]). Apart from a constant factor, $I(y)$ is the y-component H_y of the magnetic field of the diffracted wave; we have

$$I(y) = 0 \quad \text{if} \quad \tfrac{1}{2}\beta a < |y| < \tfrac{1}{2}a. \tag{3.4}$$

The other boundary conditions are

$$E_z(0, y) = -E_0 \cos\frac{\pi y}{a} \qquad \text{for} \qquad -\frac{1}{2}\beta a < y < \frac{1}{2}\beta a \tag{3.5}$$

$$\lim_{|x|\to\infty} \left| E_z(x, y) - C \cos\frac{\pi y}{a} \exp\{-i\,|x|\,k_{1/2}\} \right| = 0 \tag{3.6}$$

for a suitably chosen constant C. This is not the radiation condition of Sommerfeld which, in general, cannot be satisfied in a wave guide, not even in the form allowing the boundaries to extend to the infinite parts of the space, (see Rellich [9]).

The condition of the finiteness of the total energy in a finite part of the space gives

$$\sum_{m=1}^{\infty} |A_m|^2 \, |k_{m+1/2}|^{-1} < \infty. \tag{3.7}$$

This can be shown by integrating the square of the y-component of the magnetic field over the volume. As a consequence of (3.7) we see that the series in (3.2) can be used for the representation of E_z if $x = 0$. As in §1, we can show that

$$\frac{\bar{E}_0 k_{1/2}}{60\pi k} A_0 = A_0\bar{A}_0 + ik_{1/2} \sum_{m=1}^{\infty} A_m\bar{A}_m / |k_{m+1/2}|. \tag{3.8}$$

This proves the *uniqueness theorem.* If the A_m satisfy (3.7) they are uniquely determined by (3.4), (3.5). The complete solution is also uniquely determined because of (3.6).

The reflection coefficient becomes

$$R = -\frac{60\pi k}{E_0 k_{1/2}} A_0 . \tag{3.9}$$

It can be shown that $|R|$ is always different from 0 and 1 if $\beta \neq 0$, $\beta \neq 1$. We can also prove an inequality which $|R|$ must satisfy.
Since

$$\int_{-\frac{1}{2}a}^{\frac{1}{2}a} |E_z(0, y)|^2 \, dy \geqq \int_{-\frac{1}{2}\beta a}^{\frac{1}{2}\beta a} |E_0|^2 \left(\cos\frac{\pi y}{a}\right)^2 dy \tag{3.10}$$

it can be shown (from (3.8), (3.2)) that

$$|R|^2 \geqq \beta + \frac{1}{\pi}\sin \pi\beta - \frac{1}{2}\left(1 - \frac{1}{4}\mu^2\right)^{1/2}, \tag{3.11}$$

where

$$\mu = \lambda/a; \qquad 2 > \mu > 2/3. \tag{3.12}$$

4. *The Linear Equations for* $\beta = \frac{1}{2}$

We normalize E_0 in such a way that

$$\frac{60\pi k}{E_0 k_{1/2}} = 1 \tag{4.1}$$

and we introduce $\zeta = \pi y/a$ as a new variable. The functions

$$\phi_m(\zeta) = \sqrt{2}\cos((2m + 1)\pi\zeta) \tag{4.2}$$

form a complete orthonormal set of even functions in $-\frac{1}{2} \leqq \zeta \leqq \frac{1}{2}$. The equations to be satisfied are (if $\beta = \frac{1}{2}$):

$$\sum_{m=0}^{\infty} A_m\phi_m(\zeta) = 0 \qquad \text{if} \qquad +\frac{1}{4} < |\zeta| < \frac{1}{2} \tag{4.3}$$

$$\sum_{m=0}^{\infty} A_m \frac{k_{1/2}}{k_{m+1/2}}\phi_m(\zeta) = \phi_0(\zeta) \qquad \text{if} \qquad -\frac{1}{4} < \zeta < \frac{1}{4}. \tag{4.4}$$

According to Meixner [7] we shall have to expect that the series in (4.3) becomes infinite like $(-\zeta^2 + 1/16)^{-1/2}$ in $(-\frac{1}{4}, \frac{1}{4})$; from (A.19) in the Appendix we see that in this case

$$A_{2l} = C_1 l^{-1/2} + \mathcal{O}(l^{-1}), \qquad l = 1, 2, 3, \cdots \tag{4.5}$$

$$|A_{2r+1}r| \leqq C_2, \qquad r = 0, 1, 2, \cdots \tag{4.6}$$

where C_1, C_2 are constants.

From H. L. Schmid's lemma (Appendix I) and from (4.3), (4.4) we find the linear equations

$$x_n + \theta_n \sum_{m=0}^{\infty} s_{n,m}x_m = \delta_{0,n} \tag{4.7}$$

where

$$\theta_n = \frac{k_{1/2} - k_{n+1/2}}{k_{1/2} + k_{n+1/2}} \tag{4.8}$$

$$= -\frac{1 - \dfrac{2ik_{1/2}}{\mu(n + 1/2)}\left(1 - \mu^{-2}\left(n + \dfrac{1}{2}\right)^{-2}\right)^{1/2} - \dfrac{2 - \mu^2/4}{\mu^2(n + 1/2)^2}}{1 - (2n + 1)^{-2}}$$

(4.9) $$s_{n,m} = 2 \int_{-1/4}^{1/4} \phi_n(\zeta)\phi_m(\zeta)\, d\zeta - \delta_{n,m}$$

and where the original unknown quantities A_n are given by

(4.10) $$A_n = \frac{1}{2} \sum_{m=0}^{\infty} (\delta_{n,m} + s_{n,m})x_m .$$

Introducing the new variables

$$(-1)^l x_{2l} = \gamma_l , \qquad (-1)^r x_{2r+1} = \eta_r , \qquad l, r = 0, 1, 2, \cdots$$

and evaluating the $s_{n,m}$ from (4.9) we find

(4.11) $$\gamma_0 = 1$$

(4.12) $$\gamma_l + \theta_{2l}\pi^{-1} \sum_{r=0}^{\infty} \frac{\gamma_r}{l + r + \frac{1}{2}} + \theta_{2l}\pi^{-1} \sum_{r=0}^{\infty} \frac{\eta_r}{-l + r + \frac{1}{2}} = 0, \qquad l > 0$$

(4.13) $$\eta_r + \theta_{2r+1}\pi^{-1} \sum_{l=0}^{\infty} \frac{\gamma_l}{-l + r + \frac{1}{2}} - \theta_{2r+1}\pi^{-1} \sum_{l=0}^{\infty} \frac{\eta_l}{l + r + \frac{3}{2}} = 0, \quad r \geqq 0.$$

We can use (4.13) in order to express the γ_l in terms of the η_r . For this purpose we have to apply the Linfoot-Shepherd inversion (cf. Appendix I) in the case where the value of the parameter θ in (A.10) is $-\frac{1}{2}$. The inversion formula is not unique in this case, but the γ_l are uniquely determined if we apply (A.10) in a formal way, as will be seen later. That this is permitted could also be shown by a closer analysis of (4.12), (4.13). We can eliminate the γ_l by using (4.11), (4.12). The result is the following set of linear equations for the η_r in which we did not yet neglect any terms whatsoever:

(4.14) $$\sum_{r=0}^{\infty} \frac{(\frac{1}{2})_r}{r!} \left\{ -\theta_{2r+1}^{-1} + \frac{r + \frac{1}{2}}{r + 1} \right\} \eta_r = 1$$

(4.15) $$\sum_{r=0}^{\infty} \frac{(\frac{1}{2})_r}{r!} \left\{ \frac{\theta_{2l} - \theta_{2r+1}^{-1}}{-l + r + \frac{1}{2}} + \frac{1 - \theta_{2l}\theta_{2r+1}^{-1}}{l + r + 1} \right\} \left(r + \frac{1}{2}\right) \eta_r = 0, \; l = 1, 2, 3, \cdots .$$

We expand the θ_m in (4.8) in a series of powers of m^{-1}, $m = 1, 2, 3, \cdots$, neglecting all terms of order m^{-2} and of higher order. Substituting these approximations for the θ_{2l} , θ_{2r+1} in (4.14), (4.15) we obtain the *approximate equations*:

(4.16) $$\sum_{r=0}^{\infty} \frac{(\frac{1}{2})_r}{r!} \left(r + \frac{1}{2}\right) \left\{ \frac{1}{r + \frac{1}{2}} + \frac{1}{r + 1} + \frac{2\sigma}{(r + \frac{1}{2})(r + \frac{3}{4})} \right\} \eta_r = 1$$

(4.17) $$\sum_{r=0}^{\infty} \frac{(\frac{1}{2})_r}{r!} \left(r + \frac{1}{2}\right) \left\{ \frac{1}{-l + r + \frac{1}{2}} + \frac{1}{l + r + 1} - \frac{1}{r + \frac{3}{4}} \right\} \eta_r = 0,$$

$$l = 1, 2, 3, \cdots .$$

where

$$\sigma = \frac{1}{2}\frac{i(1 - \mu^2/4)^{1/2}}{\mu}, \qquad 2 > \mu > \frac{2}{3}. \tag{4.18}$$

We now apply the inversion formula (A.14) in the appendix to (4.16), (4.17); the result is

$$\eta_r + \frac{(\frac{1}{2})_r}{r!}\sum_{s=0}^{\infty}\eta_s\frac{(\frac{1}{2})_s(s+\frac{1}{2})}{s!}\left[\frac{\sigma}{(s+\frac{1}{2})(s+\frac{3}{4})} + \frac{\frac{1}{2}}{s+\frac{3}{4}}\right] = \frac{1}{2}\frac{(\frac{1}{2})_r}{(r+1)!}, \tag{4.19}$$

$$r = 0, 1, 2, \cdots .$$

In deriving (4.19) we have used the formulas (A.20) to (A.22). By a repeated application of these formulas and by putting

$$\eta_r = C\frac{(\frac{1}{2})_r}{(r+1)!} \tag{4.20}$$

we find for the constant C the value

$$C = \frac{1}{2}\left\{1 + 16\sigma/\pi + \left(\frac{1}{4} - 2\sigma\right)\left[\Gamma\left(\frac{3}{4}\right)/\Gamma\left(\frac{5}{4}\right)\right]^2\right\}^{-1}. \tag{4.21}$$

In order to determine the γ_l, we us again (4.13) to express the γ_l in terms of η_r. Combining this with (4.12), we find

$$\mathbf{g} = -A^{-1}\left(-\frac{1}{2}\right)\left[I - \frac{1}{\pi}H\left(\frac{3}{2}\right)\right]\mathbf{h} \tag{4.22}$$

where the matrices $A(-\frac{1}{2})$, I, $H(\frac{3}{2})$ are defined as in the Appendix (A.6), and $\mathbf{g}$, $\mathbf{h}$, denote the vectors with the components

$$(1 + \theta_{2l}^{-1})\gamma_l\,, \qquad (1 + \theta_{2r+1}^{-1})\eta_r\,. \tag{4.23}$$

If we substitute again

$$-1 - 4\sigma/(m + \tfrac{1}{2}), \qquad \sigma = \tfrac{1}{2}i\mu^{-1}(1 - \mu^2/4)^{1/2} \tag{4.24}$$

for θ_m^{-1}, $m = 1, 2, 3, \cdots$, we obtain from (4.22), (4.20) and from (A.22)

$$\gamma_l = C\left\{\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{5}{4})}\right\}^2\frac{(\frac{1}{2})_l}{l!}, \qquad l = 1, 2, 3, \cdots . \tag{4.25}$$

Now we can compute the coefficients A_n from (4.10), (4.20), (4.21), (4.25), (4.11), (4.12), (4.13). The result is

Theorem 2. *If we substitute* $-1 \pm 4\sigma m^{-1}$ *for* $\theta_m^{\pm 1}$ *in* (4.12), (4.13), (4.22), *we obtain the following approximate values for the constants* A_n *which determine the diffracted wave*:

(4.26) $$A_0 = \frac{1}{2} + \frac{1}{\pi} + \frac{\pi - 2}{2\pi} \frac{1 + 2Q_1}{1 + Q_1 + \sigma Q_2}$$

(4.27) $$A_{2l} = (-1)^l (1 - \theta_{2l}^{-1}) \frac{Q_1}{1 + Q_1 + \sigma Q_2} \frac{(\frac{1}{2})_l}{l!}, \qquad l = 1, 2, \cdots$$

(4.28) $$A_{2r+1} = (-1)^r (1 - \theta_{2r+1}^{-1}) \frac{\frac{1}{4}}{1 + Q_1 + \sigma Q_2} \frac{(\frac{1}{2})_r}{(r + 1)!}, \qquad r = 0, 1, \cdots$$

where

(4.29) $$Q_1 = \left\{\frac{1}{2} \Gamma\left(\frac{3}{4}\right)/\Gamma\left(\frac{5}{4}\right)\right\}^2, \qquad Q_2 = \frac{16}{\pi} - 8Q_1$$

(4.30) $$\sigma = \tfrac{1}{2} i\mu^{-1}(1 - \mu^2/4)^{1/2}, \qquad \mu = \lambda/a, \qquad \tfrac{2}{3} < \mu < 2$$

and where the θ_m , $m = 1, 2, 3, \cdots$, *are given by* (4.8).

The A_n satisfy (3.7).

Although A_0 (which now is the reflection coefficient R) satisfies the inequality (3.11), the approximation (4.26) is unsatisfactory since $A_0 \to 1.07$ (instead of $A_0 \to 1$) if $\sigma \to 0$. This is due to the fact that the expansion of θ_m in a series of powers of m^{-1} is not also an expansion in a series of powers of σ. Therefore the higher terms of this expansion would contribute to the terms of A_0 which do not involve σ.

5. *Concluding Remarks*

(i) *Higher Approximations.* The formulas of theorems 1 and 2 may be characterized as a second and a first approximation respectively, according to the number of terms in the expansion of the τ_m , θ_m which have been kept for the final setup of the linear equations. If we wish to deal with higher approximations, it is not necessary to develop any new methods. It can be shown that the solution of the original (exact) system of linear equations involves the inversion of a bounded linear operator T at a point of its spectrum which lies on the boundary of the spectrum of T^*T (where T^* denotes the adjoined operator). The Linfoot-Shepherd inversion for $\theta = \pm\frac{1}{2}$ is an inversion of this type. After its application the solution of an approximation of finite order of the given system of linear equations requires only the inversion of linear operators which have a bounded inverse and the solution of a finite system of linear equations. In the case dealt with in §2 it can be shown that (at least for a sufficiently large μ) the solutions for the n-th approximation tend towards the solution of the exact system of linear equations. By a different method, Lamb [2] has obtained a first approximation for the solution of the problem of §2. However, it seems that his method does not lead to an approximation of a higher degree.

(ii) *The Restriction* $\beta = \frac{1}{2}$. It can be shown that the Linfoot-Shepherd inversion with $\theta = \frac{1}{2}$ is equivalent to the solution of the first boundary value problem of Laplace's equation in two dimensions, where the boundary consists of all the intervals

$$(5.1) \qquad x = 0, \qquad n - \tfrac{1}{4} \leqq y \leqq n + \tfrac{1}{4}, \qquad n = 0, \pm 1, \pm 2, \cdots$$

of a straight line (the y-axis) and where the boundary conditions are periodic and of period one. This explains why the Linfoot-Shepherd formula can be applied to the problem of §2, which, for $\mu \to \infty$, reduces to a problem connected with Laplace's equation. It is to be expected that a solution of the corresponding boundary value problem for $\beta \neq \frac{1}{2}$ will lead to the right generalization of this inversion formula.

(iii) *The Case of a Diffracting Thin Wire.* We can replace the diffracting strip by a thin wire of radius ρ, the length of which equals the width of the strip. In this case the coefficients of the linear equations (2.4) or (4.12), (4.13) must be changed in the following way:

In (2.4) substitute

$$(5.2) \qquad \tfrac{1}{2}k_m^2 J_0(k_m\rho)H_0^{(2)}(k_m\rho), \qquad m = 0, 1, 2, \cdots; \qquad k_0 = k,$$

for k_m, where J_0, $H_0^{(2)}$ denote Bessel functions of the first and third kind respectively.

Instead of (4.12), (4.13) use again (2.4) and replace k_m by

$$(5.3) \qquad \tfrac{1}{2}k_m^2 J_0(k_m\rho) S_m(\rho, a, d)$$

where d denotes the y-coordinate of the axis of the wire and where

$$(5.4) \quad S_m = H_0^{(2)}(k_m\rho) + 2 \sum_{n=1}^{\infty} H_0^{(2)}(2ank_m) - \sum_{n=-\infty}^{\infty} H_0^{(2)}(|(2n+1)a + 2d| k_m).$$

The quantities which are to be substituted for the k_m have the same asymptotic behavior as the k_m for $m \to \infty$, and therefore a first approximation for the linear equations can be derived and dealt with in the same way as in the case of a diffracting strip. For the uniqueness theorem cf. Schaefke [10]. The S_m have been tabulated, because they play a role for the computation of the effect of a diffracting wire which connects the opposite boundaries of a wave guide.

Appendix

1. *H. L. Schmid's Lemma*

Let $\phi_m(x)$, $m = 0, 1, 2, \cdots$, be a complete set of orthonormal functions for the interval (0, 1). Let β be a real number, $0 < \beta < 1$ and let

$$\sum_{m=0}^{\infty} A_m \phi_m(x) = 0 \quad \text{if} \quad \beta < x < 1 \tag{A.1}$$

$$\sum \kappa_m A_m \phi_m(x) = \phi_0(x) \quad \text{if} \quad 0 < x < \beta. \tag{A.2}$$

We define

$$u_{n,m} = \int_0^\beta \phi_n(x)\phi_m(x)\, dx; \qquad v_{n,m} = \int_\beta^1 \phi_n(x)\phi_m(x)\, dx. \tag{A.3}$$

Then the matrices U, V with the general element $u_{n,m}$, $v_{n,m}$ satisfy

$$UV = VU = 0; \qquad U + V = I; \qquad U^2 = U, \qquad V^2 = V, \tag{A.4}$$

where I denotes the identity. If we denote by $\mathbf{a}$ the vector with the components A_n, then there exists a vector $\mathbf{x}$ such that

$$\mathbf{a} = U\mathbf{x}, \qquad (V + KU)\mathbf{x} = \mathbf{e} \tag{A.5}$$

where $\mathbf{e} = (1, 0, 0, \cdots)$ and where K denotes the diagonal matrix $(\delta_{n,m}\kappa_m)$. This statement is valid if the functions on the left hand side of (A.1), (A.2) are absolutely integrable and if it is permitted to multiply (A.1), (A.2) by $\phi_n(x)$ and to integrate the left hand side term by term in $(0, \beta)$ and $(\beta, 1)$. A proof can be derived from Schaefke [10].

2. *Inversion Formulas*

Let θ be a real parameter, $\theta \neq 0, -1, -2, \cdots$ and let $A(\theta)$, $H(\theta)$ be the matrices with the general elements

$$\begin{aligned} a_{n,m}(\theta) &= \frac{\sin \pi\theta}{\pi} \frac{1}{-n + m + \theta}, \\ h_{n,m}(\theta) &= \frac{1}{n + m + \theta}, \qquad n, m = 0, 1, 2, \cdots \end{aligned} \tag{A.6}$$

where n denotes the row and m denotes the column of the matrix elements $a_{n,m}$, $h_{n,m}$. The matrices $A(\theta)$, $H(\theta)$ are bounded; cf. [11]. The following is a special case of *Titchmarsh's inversion formula* [13]: The matrix

$$T = 2^{-1/2} A\left(\frac{1}{2}\right) + \frac{1}{\pi\sqrt{2}} H\left(\frac{1}{2}\right) \tag{A.7}$$

has a bounded inverse T^{-1} the general element of which is

$$\frac{1}{\pi\sqrt{2}} \left\{ \frac{\epsilon_n}{-n + m + \frac{1}{2}} + \frac{\epsilon_n}{n + m + \frac{1}{2}} \right\}; \qquad \epsilon_0 = 1, \qquad \epsilon_1 = \epsilon_2 = \cdots = 2. \tag{A.8}$$

Titchmarsh has shown, that under very wide conditions for a vector $\mathbf{y}$

$$T^{-1}\mathbf{y} = \mathbf{x} \quad \text{if} \quad T\mathbf{x} = \mathbf{y} \tag{A.9}$$

and vice versa.

The Linfoot-Shepherd inversion formula gives a formal inverse $A^{-1}(\theta)$ of $A(\theta)$. The general element of $A^{-1}(\theta)$ is

$$-\frac{\sin \pi\theta}{\pi} \frac{\Gamma(1 + n + \theta)}{n!} \frac{1}{-n + m - \theta} \frac{\Gamma(1 + m - \theta)}{m!}, \; n, m = 0, 1, \cdots. \tag{A.10}$$

Linfoot and Shepherd [3] stated certain sufficient conditions for a vector $\mathbf{y}$ such that

$$A^{-1}(\theta)\mathbf{y} = \mathbf{x} \quad \text{if} \quad A(\theta)\mathbf{x} = \mathbf{y}. \tag{A.11}$$

They also showed that for $\theta \geqq 0$ the homogeneous equations

$$A(\theta)\mathbf{x} = 0 \tag{A.12}$$

do not have any solution whatsoever except $\mathbf{x} = 0$ and that for $-1 < \theta < 0$, (A.12) has the only non-trivial solution $\mathbf{x} = \{x_m\}$ where

$$x_m = C\frac{(\theta + 1)_m}{m!}.$$

Here C is an arbitrary constant and

$$\begin{aligned} (u)_m &= \frac{\Gamma(u + m)}{\Gamma(u)} \\ &= u(u + 1) \cdots (u + m - 1) \quad \text{if} \quad m = 1, 2, \cdots; \quad (u)_0 = 1. \end{aligned} \tag{A.13}$$

It can be shown that $A^{-1}(\theta)$ is bounded if and only if $-\frac{1}{2} < \theta < \frac{1}{2}$ cf. [5], [10]. A result connected with (A.10) is:

The inverse of $\pi A(\frac{1}{2}) + H(1) = G$ is a bounded matrix $G^{-1} = (g_{n,m})$ the general element of which is

$$g_{n,m} = -\left\{\frac{(n + \frac{1}{2})(\frac{1}{2})_n}{n!} \frac{(\frac{1}{2})_m}{m!}\right\}^2 \left\{\frac{1}{-n + m + \frac{1}{2}} + \frac{1}{n + m + 1}\right\}. \tag{A.14}$$

The boundedness of G^{-1} follows from (A.14) because

$$\begin{aligned} &\frac{1}{-n + m + \frac{1}{2}} + \frac{1}{n + m + 1} \\ &= -\frac{2m + \frac{1}{2}}{2n + \frac{3}{4}}\left\{\frac{-1}{-n + m - \frac{1}{2}} + \frac{1}{n + m + 1}\right\} \end{aligned} \tag{A.15}$$

and

$$\lim_{n\to\infty} \frac{(\frac{1}{2})_n}{n!} \sqrt{n} = \pi^{-1/2}. \tag{A.16}$$

This shows that the elements of G^{-1} can be obtained from those of $H(1) -$

$\pi A(-\frac{1}{2})$ by multiplying them by certain bounded positive factors; the rest follows from a criterion of I. Schur [11]. The proof of (A.14) can be obtained from [5]. It is not difficult to prove that G^*G has a minimum >0 and that therefore G has a bounded inverse. But although (A.14) is a formal inverse of G it is still necessary to prove that it is bounded; cf. [5].

3. *Some Special Fourier Series*

From the Hansen-Bessel integral representation for the Bessel function J_0 of the first kind

$$J_0(z) = \pi^{-1} \int_{-1}^{1} \frac{\cos zt}{(1-t^2)^{1/2}}\, dt \tag{A.17}$$

we find that

$$\sum_{m=0}^{\infty} \epsilon_m \pi J_0(\pi m\beta) \cos 2\pi mx = \begin{cases} 0 & \text{if} \quad \frac{1}{2}\beta < x < \frac{1}{2} \\ (\frac{1}{4}\beta^2 - x^2)^{-1/2} & \text{if} \quad -\frac{1}{2}\beta < x < \frac{1}{2}\beta. \end{cases} \tag{A.18}$$

The convergence of the series on the left hand side in (A.18) can be proved by expressing the partial sums as an integral which can be derived from (A.17). The behavior of the coefficients in (A.18) if $m \to \infty$ is given by

$$J_0(m\pi\beta) = \pi^{-1}\left(\frac{2}{m\beta}\right)^{1/2} \cos\left(m\pi\beta - \frac{1}{4}\right)[1 + O(m^{-1})].$$

The formula

$$\sum_{n=0}^{\infty} (-1)^n \frac{(\frac{1}{2})_n}{n!} \cos\{\pi(4n+1)x\} = \begin{cases} (2\cos 2\pi x)^{-1/2} & \text{if} \quad -\frac{1}{4} < x < \frac{1}{4} \\ 0 & \text{if} \quad \frac{1}{4} < |x| < \frac{1}{2} \end{cases} \tag{A.19}$$

can be proved by expressing the series in terms of two hypergeometric functions of argument $\exp\{\pm i4\pi x\}$.

4. *Sums*

The summations which are involved in the multiplication of infinite matrices in sections 2 and 4 can be carried out explicitly by using the following formulas and their derivatives with respect to z:

$$\frac{\Gamma(\theta)\Gamma(z)}{\Gamma(z+\theta)} = \sum_{n=0}^{\infty} \frac{(1-\theta)_n}{n!} \frac{1}{z+n}, \qquad (\Re e\,\theta > 0) \tag{A.20}$$

$$\Gamma(\theta)\left\{\frac{\Gamma(z)}{\Gamma(z+\theta)} - \frac{\Gamma(\zeta)}{\Gamma(\zeta+\theta)}\right\} = \sum_{n=0}^{\infty} \frac{(1-\theta)_n}{n!}\left\{\frac{1}{z+n} - \frac{1}{\zeta+n}\right\}, \qquad (\Re e\,\theta > -1) \tag{A.21}$$

$$\frac{\Gamma(1-2b+2a)\Gamma(1-2b)}{\Gamma(2a)} \frac{\Gamma(a-z)\Gamma(a+z)}{\Gamma(1+a-2b-z)\Gamma(1+a-2b+z)}$$

(A.22)

$$= \sum_{n=0}^{\infty} \frac{(2a)_n(2b)_n}{(1-2b+2a)_n n!} \left\{\frac{1}{a+n-z} + \frac{1}{a+n+z}\right\}; \quad (\Re b < \tfrac{1}{2}).$$

Equations (A.20), (A.21) follow from the expansion of Euler's Beta-integral in an infinite series. For a proof of (A.22) cf. [5]. The sums S_n in (2.22) have the generating function

$$\frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}+z)}{\Gamma(z+1)} = \sum_{n=0}^{\infty} (-1)^n S_{n+1} z^n \tag{A.23}$$

and the first three of them can be expressed in terms of π and log 2 by using well known properties of the Gamma Function.

Bibliography

1. C. J. Bouwkamp, *On the freely vibrating circular disc and diffraction by circular discs and apertures*, Physica, Volume 16, 1950, pp. 1–16.
2. H. Lamb, *On the reflection and transmission of electric waves by a metallic grating*, Proceedings of the London Mathematical Society, Volume 29, 1898, pp. 523–544.
3. E. H. Linfoot and W. M. Shepherd, *On a set of linear equations (II)*, Quarterly Journal of Mathematics, (Oxford Series), Volume 10, 1939, pp. 84–98.
4. W. Magnus, *Über Eindeutigkeitsfragen bei einer Randwertaufgabe von $\Delta u + k^2 u = 0$*, Jahresbericht der Deutschen mathematiker Vereinigung, Volume 52, 1943, pp. 177–188.
5. W. Magnus, *Über einige beschränkte Matrizen*, to appear in Archiv der Mathematik.
6. A. W. Maue, *Zur Formulierung eines allgemeinen Beugungsproblems durch eine Integralgleichung*, Zeitschrift für Physik, Volume 126, 1949, pp. 601–618.
7. J. Meixner, *Die Kantenbedingung in der Theorie der Beugung elektromagnetischer Wellen an vollkommen leitenden Schirmen*, Annalen der Physik (Series 6), Volume 6, 1949, pp. 1–7.
8. F. Oberhettinger and W. Magnus, *Über einige Randwertprobleme der Schwingungsgleichung*, Journal für die reine und angewandte Mathematik, Volume 186, 1945, pp. 1–9.
9. F. Rellich, *Über das asymptotische Verhalten von Lösungen von $\Delta u + \lambda u = 0$ in unendlichen Gebieten*, Jahresbericht der Deutschen mathematiker Vereinigung, Volume 53, 1943, pp. 157–165.
10. F. W. Schaefke, *Über einige unendliche lineare Gleichungssysteme*, Mathematische Nachrichten, Volume 3, 1949, pp. 40–58.
11. I. Schur, *Bemerkungen zur Theorie der beschränkten Bilinearformen*, Journal für die reine und angewandte Mathematik, Volume 140, 1911, pp. 1–28.
12. A. Sommerfeld, *Die Greensche Funktion der Schwingungsgleichung*, Jahresbericht der Deutschen mathematiker Vereinigung, Volume 21, 1912, pp. 309–353.
13. E. C. Titchmarsh, *A series inversion formula*, Proceedings of the London Mathematical Society (Series 2), Volume 26, 1927, pp. 1–11.

Wiener-Hopf Techniques and Mixed Boundary Value Problems*

By S. N. KARP
New York University

Introduction

J. Schwinger [1] has shown how Green's theorem may be utilized to formulate certain of the boundary value problems of electromagnetic theory and acoustics relating to the equation $\Delta u + k^2u = 0$, as Wiener-Hopf integral equations, the boundaries in question being, in general, semi-infinite planes or cylinders. The Wiener-Hopf equation $f(x) = \int_0^\infty g(x_0)K(x - x_0)\, dx_0 : x > 0$, is solved by the application of the Fourier transform, together with function-theoretic considerations in the Argand plane of the transform variable. Fourier transformation results in a single equation between two unknown transforms, and this equation is solved for both unknown functions by function-theoretic techniques. The notable success of this procedure in the hands of Schwinger [1], Carlson and Heins [2], Heins [3], Levine [4], and others has stimulated further investigation. G. Carrier [5] has indeed shown that the procedure may be applied to other partial differential equations. He has illustrated this point by considering a boundary value problem for a generalized Tricomi equation,

$$u_{yy} + y^m(u_{xx} + k^2u) = 0,$$

and a semi-infinite plane obstacle. The excitation was taken to be a "plane wave," i.e., a function reducing to the latter when $m = 0$. The integral equation resulting from Green's theorem is not of Wiener-Hopf type, but when a certain Hankel transformation was applied (reducing to the Fourier transform for $m = 0$), a relation between two unknown transforms resulted, which was

---

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Sciences and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directories of the Air Force Cambridge Research Laboratories.

---

*This work was performed at Washington Square College of Arts and Science, New York University, and was supported in part by Contract No. AF-19(122)-42, with the U.S. Air Force through sponsorship of Geophysical Research Directorate, Air Force Cambridge Research Laboratories, Air Materiel Command.

solved (after an ingenious change of complex variable) by the customary techniques.[1]

In the present note the parallelism between the method of separation of variables and the Green's function integral equation method is shown to persist in the present situation as it does in problems of more classical type. This relationship leads to a characterization (from the standpoint of coordinate systems) of those problems in which the "Wiener-Hopf" type of problems (in an extended sense) arise. Certain heuristic advantages of the separation of variables procedure are also pointed out.

(*1*) To fix the ideas a simple and typical illustration of the integral equation method of Schwinger is recalled at this point. The problem relates to the diffraction of a plane wave by a semi-infinite plane; it has also been solved in the present manner by Copson [6] independently of Schwinger. The mathematical problem is to find a function $u(x, y)$ such that,

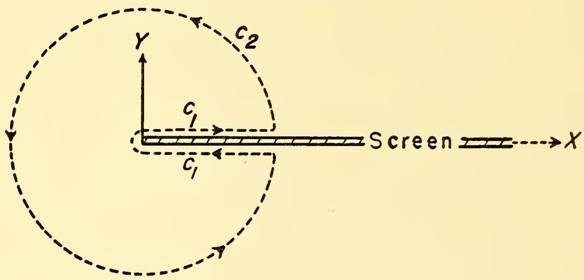


FIGURE 1

(1.1) $u_{xx} + u_{yy} + k^2u = 0, \quad \text{where} \quad k = k_1 + ik_2, \quad 1 \gg k_2 > 0$

(1.2) $$\left.\frac{\partial u}{\partial y}\right)_{y=0} = 0 \quad \text{for} \quad x > 0,$$

(1.3) $$u = u_0 + u_1,$$

(1.4) $$u_0 = \exp\{ik(x\cos\theta_0 + y\sin\theta_0)\},$$

(1.5) $$u_1 \to 0 \quad \text{as} \quad r \to \infty, \text{ off the positive } x\text{-axis},$$

(1.6) $$u \text{ regular in exterior of the positive } x\text{-axis}.$$

Here x and y are Cartesian coordinates, r and θ the usual polar coordinates, and θ_0 gives the direction of the incident field u_0.

[1]The present investigation was suggested by this work and by a private remark of this author that the transform to employ is suggested by the form of the product solution of the partial differential equation, in the form $f(x)g(y)$.

The free space Green's function is introduced. It is

$$G(x, y: x_0, y_0) = \frac{i}{4} H_0^{(1)}[k\{(x - x_0)^2 + (y - y_0)^2\}^{1/2}]$$

where $H_0^{(1)}$ is the Hankel function of the first kind. Green's theorem is applied to the functions u and G, along the contour $c = c_1 + c_2$ indicated in Figure 1; c_2 is an arbitrarily large circle.

Using the singularity of G at $(x, y) = (x_0, y_0)$, and the boundary conditions (1.2) and (1.6), and letting $c_2 \to \infty$, one finds

$$\begin{aligned} u(x, y) = \int_0^\infty [u] \frac{\partial G}{\partial y_0} (x, y: x_0, y_0)\Big]_{y_0=0} dx_0 \\ + \exp \{ik(x \cos \theta_0 + y \sin \theta_0)\} \end{aligned} \tag{1.7}$$

where $[u]$ is the discontinuity in u across the screen. Differentiating with respect to y and applying (1.2) again one finds

$$0 = ik \sin \theta_0 \exp \{ik\, x \cos \theta_0\} + \int_0^\infty [u(x_0)]K(x - x_0)\, dx_0, \qquad x > 0, \tag{1.8}$$

where

$$K(x - x_0) \quad \text{is} \quad \frac{\partial^2 G}{\partial y\, \partial y_0}\Big|_{y=y_0=0}.$$

This is a typical inhomogeneous Wiener-Hopf equation. One now defines

$$\begin{aligned} &g(x) \equiv 0, \quad x > 0 \qquad f(x_0) = [u], \quad x_0 > 0 \\ &h(x) \equiv 0, \quad x < 0 \qquad f(x_0) = 0, \quad x_0 < 0 \\ &h(x) = ik \sin \theta_0 \exp \{ik(x \cos \theta_0)\}, \quad x > 0 \end{aligned} \tag{1.9}$$

and then equation (1.8) may be rewritten,

$$g(x) = h(x) + \int_{-\infty}^{\infty} f(x_0)K(x - x_0)\, dx_0, \qquad -\infty < x < \infty \tag{1.10}$$

in terms of the two unknown functions $g(x)$ and $f(x_0)$. Multiplying by $\exp \{-i\alpha x\}$ and integrating with respect to α, one finds

$$\bar{g}(\alpha) = \bar{h}(\alpha) + \bar{f}(\alpha)\bar{K}(\alpha) \tag{1.11}$$

where for instance $\bar{K}(\alpha) = \int_{-\infty}^{\infty} K(t) \exp \{-i\alpha t\}\, dt$. This equation holds in the common strip of regularity (provided such exists) of the functions $\bar{g}(\alpha)$, $\bar{f}(\alpha)$, $\bar{K}(\alpha)$, $\bar{h}(\alpha)$ considered as functions of α.

The technique of determining the regions of regularity of these functions is

as follows. One assumes for $f(x_0)$ say, a behavior like $\exp\{i\kappa x_0\}$ for large x_0, where $\mathscr{I}m(\kappa) > 0$. Then from the formula

$$\bar{f}(\alpha) = \int_0^\infty f(x_0) \exp\{-i\alpha x_0\}\, dx_0 \tag{1.12}$$

we find that $\bar{f}(\alpha)$ will be regular in the lower half-plane $\mathscr{I}m(\alpha) < \mathscr{I}m(\kappa)$, which starts above the real axis in the α-plane. One finds that $\bar{h}(\alpha)$ has similar behavior, while $\bar{g}(\alpha)$ is regular in an upper half-plane starting below the real axis. Also

$$\bar{K}(\alpha) = \frac{i}{2}(k^2 - \alpha^2)^{1/2}, \qquad \text{regular for} \qquad |\mathscr{I}m(\alpha)| < \mathscr{I}m(k).$$

Hence in a strip enclosing the real axis, we have,

$$g_+(\alpha) = \frac{k \sin\theta_0}{\alpha - k\cos\theta_0} + \bar{f}_-(\alpha)\cdot\frac{i}{2}(k^2 - \alpha^2)^{1/2}, \tag{1.13}$$

the first term on the right being $\bar{h}_-(\alpha)$. The subscripts $+$, $-$, denote regularity in upper and lower half-planes respectively.

$\bar{K}(\alpha)$ is now factored in the form $K_-(\alpha)/K_+(\alpha)$ where $K_-(\alpha) = (k - \alpha)^{+1/2}$ and $K_+(\alpha) = (k + \alpha)^{-1/2}$. After multiplying by $K_+(\alpha)$ and applying suitable manipulations, one finds

$$\begin{aligned} &\frac{g_+(\alpha)}{(k+\alpha)^{1/2}} - \frac{k\sin\theta_0}{\alpha - k\cos\theta_0}\left[\frac{1}{(k+\alpha)^{1/2}} - \frac{1}{(k + k\cos\theta_0)^{1/2}}\right] \\ &\qquad = \frac{k\sin\theta_0}{(k + k\cos\theta_0)^{1/2}(\alpha - k\cos\theta_0)} + \frac{i}{2}\bar{f}_-(\alpha)(k - \alpha)^{1/2}, \end{aligned} \tag{1.14}$$

an equation whose left and right sides are regular in upper and lower half-planes respectively. Since these half-planes overlap, an entire function, $E(\alpha)$ say, is defined by (1.14), the left and right sides being various representations. The growth of this function at infinity is studied via the growths of $g_+(\alpha)$, and $f_-(\alpha)$, and it is found to be of negative fractional degree; hence it is zero by a modification of Liouville's theorem. Hence the left and right sides of (1.14) are identically zero, and this determines $g_+(\alpha)$ and $f_-(\alpha)$ simultaneously; for instance

$$f_-(\alpha) = \frac{2ik\sin\theta_0}{(k + k\cos\theta_0)^{1/2}}\cdot\frac{1}{(k-\alpha)^{1/2}}\cdot\frac{1}{\alpha - k\cos\theta_0} \tag{1.15}$$

The growths are studied, for the above purposes, by noting, for instance in (1.12), that as $\alpha \to \infty$ in the lower half plane the resulting exponential decay makes the integrand negligible except in the neighborhood of $x_0 = 0$. Hence if we assume $f(x_0) \sim x_0^{p-1}$ near $x_0 = 0$, we find

$$f_-(\alpha) \sim \int_0^\infty x_0^{p-1} \exp\{-i\alpha x_0\}\, dx_0 \sim \text{constant}\, \alpha^{-p},$$

(1.16)

$$\alpha \to \infty \quad \text{with} \quad \mathscr{I}m(\alpha) < 0.$$

Here, also, we require $p > 0$, for convergence near the origin.

(*2*) We now wish to emphasize a few features of the method of separation of variables. In this method a linear partial differential equation is written in a certain separable coordinate system ξ, η; then there exist solutions in the form,

$$V_\alpha = f_\alpha(\xi) g_\alpha(\eta)$$

for various values of the separation parameter α. The functions f_α , g_α arise from certain ordinary differential equations, of second order, and are combinations of their solutions, chosen to meet certain conditions of the problem in question. A more general solution is then

$$V = \int_c A(\alpha) f_\alpha(\xi) g_\alpha(\eta)\, d\alpha$$

where c is a suitable contour in the complex plane. When the contour encloses a sequence α_n of poles (only) of the integrand, one may also write

$$V = \sum A_n f_{\alpha_n}(\xi) g_{\alpha_n}(\eta),$$

a Fourier series representation. Now the classical application of this method to boundary value problems envisions a boundary identical with the curve $\xi = \xi_0$, say. The solution of a boundary value problem is written in the form $u = u_0 + V$ where u_0 is the incident field, and where the form of V has been given above. Then V has to adopt certain boundary values on $\xi = \xi_0$ for all η, these values typically serving to cancel those of u_0 , at $\xi = \xi_0$. Thus one has, for all η in the region of interest,

$$0 = F(\eta) + \int_c A(\alpha) f_\alpha(\xi_0) g_\alpha(\eta)\, d\alpha \tag{2.1}$$

with $F(\eta)$ given. This is an equation for the "Fourier coefficient" $A(\alpha)$, and may be inverted by the use of a suitable "Fourier" transform theorem. The existence of such generalized transform theorems has been discussed by Titchmarsh [7] and others.

However, one may consider cases in which the coordinate system chosen does not conform to the boundary, i.e. the boundary in question is only part of the curve $\xi = \xi_0$. This means in other words that one represents the solution in terms of the eigen-functions of a simpler and different problem. In such a coordinate system equation (2.1) will still hold in the range of interest in η, i.e., on the obstacle or boundary. But for $\xi = \xi_0$, and in other ranges of η, different conditions may be derived and will hold. These may result for instance from a requirement of regularity for u in the exterior of the obstacle. In this manner

or in other ways, a mixed problem will arise, expressible in a *pair* of integral equations, for example

$$F(\eta) = \int_{c_1} A(\alpha) f_\alpha(\xi_0) g_\alpha(\eta)\, d\alpha \tag{2.2a}$$

$$G(\eta) = \int_{c_2} A(\alpha) f'_\alpha(\xi_0) g_\alpha(\eta)\, d\alpha \tag{2.2b}$$

where $F(\eta)$ is given for η on the obstacle or boundary, and $G(\eta)$ is given for η off the boundary. Possibly also, the point set ($\xi = \xi_0$, η off the obstacle) may be further subdivided into various regions on which diverse conditions hold. In the present case the transform theorem alone is not directly applicable, for the value of the integral in (2.2a) say, is only given in part of the range of η on the boundary.

We wish to observe here that such a *two part problem* is often solvable by the intervention of function-theoretic techniques, and that if the boundary value problem for the partial differential equations were instead formulated as an integral equation with Green's function, it would give rise under these circumstances to a "Wiener-Hopf" problem in the extended sense, in terms of the transform relating to the coordinate system in which the problem is *two part.* This is in accordance with footnote 1, but is *not* restricted to Cartesian coordinates.

On the other hand, of course, if the problem is of the classical type first discussed, (and thus solvable by the use of transforms alone), then the corresponding Green's function integral equation is also directly reducible, *by use of the transform in question,* to a simple equation between transforms, which can be solved purely algebraically.

(*3*) We wish to illustrate the above remarks in concrete cases. Our first illustration relates to the problem of section *1*, Equations (1.1) to (1.6)

A solution by separation of variables which is small at $y = \pm\infty$, or alternatively, corresponds to waves diverging from the positive x axis is, in Cartesian coordinates,

$$\exp\{i\alpha x\}\cdot\exp\{i(k^2 - \alpha^2)^{1/2}\,|\,y\,|\}.$$

Hence we write, in this inappropriate coordinate system,

$$u_1 = \pm\frac{1}{2}\int_{-\infty}^{\infty} \psi(\alpha)\exp\{i[\alpha x + (k^2 - \alpha^2)^{1/2}\,|\,y\,|]\}\, d\alpha \tag{3.1}$$

where the plus or minus sign refers to $y > 0$ or $y < 0$ respectively. Then $\partial u_1/\partial y$ is continuous at $y = 0$, while u_1 may be discontinuous. Using (1.6), we find, with $u = u_0 + u_1$

$$\int_{-\infty}^{\infty} \psi(\alpha)\exp\{i\alpha x\}\, d\alpha = 0, \qquad \text{for} \qquad x < 0 \tag{3.2a}$$

$$ik \sin \theta_0 \exp \{ik(x \cos \theta_0)\} + \frac{i}{2} \int_{-\infty}^{\infty} \psi(\alpha)(k^2 - \alpha^2)^{1/2} \exp \{i\alpha x\}\, d\alpha = 0, \tag{3.2b}$$

$$\text{for} \quad x > 0,$$

cf. (2.2a), (2.2b). We now transform the integrals in these equations into contour integrals by adding infinite semicircles. The contributions of the latter are to be vanishingly small. In (3.2a), for instance, this is the case when the semicircle is in the lower half-plane, (in virtue of the decay of the exponential there for $x < 0$), provided $\psi(\alpha)$ is merely of algebraic growth at ∞ in that half-plane. The upper half-plane is suitable for equation (3.2b). Expressing the excitation in suitable form for $x > 0$, by use of the Fourier theorem, we find

$$\int_{C_1} \psi(\alpha) \exp \{i\alpha x\}\, d\alpha = 0, \tag{3.3a}$$

$$\int_{C_2} \left(\frac{i}{2} \psi(\alpha)(k^2 - \alpha^2)^{1/2} + \frac{k \sin \theta_0}{\alpha - k \cos \theta_0}\right) \exp \{i\alpha x\}\, d\alpha = 0, \tag{3.3b}$$

where C_1, C_2 are the semicircles in the lower and upper half-planes, which have been referred to above. By Cauchy's theorem, (3.3a) would be fulfilled if $\psi(\alpha)$ is regular in the closed lower half-plane. Define

$$\phi(\alpha) \equiv \frac{i}{2} \psi(\alpha)(k^2 - \alpha^2)^{1/2} + \frac{k \sin \theta_0}{\alpha - k \cos \theta_0}. \tag{3.4}$$

Then (3.3b) would be fulfilled if $\phi(\alpha)$ is regular in the closed upper half-plane. Thus in (3.4) we may write $\phi = \phi_+$, $\psi = \psi_-$, and we are in the position given by (1.13), with ϕ corresponding to g, ψ corresponding to f. We see here the role played by the algebraic growth and the region of regularity of $\psi(\alpha)$ and $\phi(\alpha)$.

We wish to add here a remark on uniqueness. Due to the integrability at the origin required by the methods of section (*1*), it was required there that

$$g(x) \sim (-x)^{q-1}, \qquad f(x_0) \sim x_0^{p-1}, \qquad q > 0, \qquad p > 0,$$

in the notation of that section. This affected the growth of $g_+(\alpha)$, $f_-(\alpha)$, and ensured the vanishing of the entire function $E(\alpha)$. However, the procedure of section *1* is not obligatory. From the present heuristic standpoint, after manipulating (3.4) in the manner of section (*1*), we find, cf. (1.14)

$$E(\alpha) = \frac{\phi_+(\alpha)}{(k + \alpha)^{1/2}} - \frac{k \sin \theta_0}{\alpha - k \cos \theta_0}\left(\frac{1}{(k + \alpha)^{1/2}} - \frac{1}{(k + k \cos \theta_0)^{1/2}}\right)$$

$$= \frac{k \sin \theta_0}{(\alpha - k \cos \theta_0)(k + k \cos \theta_0)^{1/2}} + \frac{i}{2} \psi_-(\alpha)(k - \alpha)^{1/2},$$

but $E(\alpha)$ need not be zero, for certain growths of $\psi_-(\alpha)$, $\phi_+(\alpha)$. Then we have

$$(3.5) \qquad \psi_-(\alpha) = \frac{2ik \sin \theta_0}{(\alpha - k \cos \theta_0)(k - \alpha)^{1/2}(k + k \cos \theta_0)^{1/2}} + \frac{i}{2} \frac{E(\alpha)}{(k - \alpha)^{1/2}}.$$

The first term corresponds to (1.15). Take for example $\phi_+(\alpha) \sim \alpha^{1/2}$. Since ϕ corresponds to g of section (*1*), this implies $g(x) \sim (-x)^{-3/2}$ near the origin, so that it would have been excluded. The corresponding $E(\alpha)$ is then a constant in (3.5), so that a term of the form $(k - \alpha)^{-1/2}$ is added to $\psi_-(\alpha)$, or a term of the form

$$\text{signum } y \cdot \int_{-\infty}^{\infty} \frac{\exp \{i[\alpha x + (k^2 - \alpha^2)^{1/2} \mid y \mid]\}}{(k - \alpha)^{1/2}} \, d\alpha$$

is added to u_1. When $y = 0$ this is zero for negative x and a multiple of $\exp \{ikx\}/x^{1/2}$ for positive x, while its derivative with respect to y vanishes for $y = 0$, $x > 0$. Thus it corresponds to the addition of the solution

$$H_{1/2}(kr) \cos (\theta/2),$$

which does not affect the conditions of the problem. If for example $E(\alpha) = \alpha$, then we can interpret the contribution operationally as $\partial/\partial x$ of the above homogeneous singular solution. Also, the use of a suitable polynomial as $E(\alpha)$, will produce multiples of the derivatives with respect to x of the original solution u. All such solutions are clearly admissible, as has been pointed out by Bouwkamp [8], so that for uniqueness it is necessary to specify the behaviour of the solution at the origin also.

(*4*) From the present standpoint, we see that the Fourier transform solution of the integral equation (1.10) is parallel to the representation of the solution of equations (1.1) to (1.6) in Cartesian coordinates (x, y). Here the curve $\xi = \xi_0$ cf. section 2, especially equations (2.2a) and (2.2b) is the line $y = 0$; only part of this line, i.e. $x > 0$, is the boundary in question, i.e. the diffracting screen, so that we have a two-part problem. A more suitable coordinate system would be the system of polar coordinates r, θ, where the upper and lower surfaces of the screen are represented by $\theta = 0$, and $\theta = 2\pi$, respectively. In terms of this coordinate system, our boundary value problem is a one part problem of classical type. The eigen-functions are

$$J_\nu(kr) \begin{Bmatrix} \cos \nu\theta \\ \sin \nu\theta \end{Bmatrix},$$

and a more general solution is, say,

$$(4.1) \qquad u_1 = \int_{-i\infty}^{i\infty} J_\nu(kr)[A(\nu) \cos \nu\theta + B(\nu) \sin \nu\theta] \, d\nu.$$

The problem may be solved directly in this manner, without the intervention

of function theoretic techniques, when one is in possession of an inversion formula for (4.1). Such an inversion formula has been obtained by N. N. Lebedev, and employed by the latter and M. J. Kontorowich [9] in the solution of this problem. The formula reads

$$\phi(kr) = -\frac{1}{2}\int_{-i\infty}^{i\infty} \nu\bar{\phi}(\nu)J_\nu(kr)\,d\nu, \tag{4.2}$$

$$\bar{\phi}(\nu) = \int_0^\infty \phi(kr)H_\nu^{(2)}(kr)\,\frac{dr}{r}. \tag{4.3}$$

(Actually the transformation (4.3) was applied to the partial differential equation after certain preparatory modifications relating to integrability.)

When the integral equation (1.10) is considered ab initio there is nothing to suggest the use of any particular type of transform. From the present standpoint it is seen that while it is a Wiener-Hopf equation, as far as the Fourier transform is concerned, it may also be inverted directly by the use of the Bessel transform of (4.2) and (4.3). It is convenient to represent the kernel in the form

$$\frac{\text{constant}}{x\,x_0}\cdot\int_{-i\infty}^{i\infty} \lambda^2 \cot \pi\lambda H_\lambda^{(2)}(kx_0)J_\lambda(kx)\,d\lambda$$

which may be obtained by replacing the series for $H_0[k\{(x - x_0)^2 + (y - y_0)^2\}^{1/2}]$ by a contour integral, after suitable differentiations. Equation (1.10) takes the form

$$\begin{aligned} &ikx \sin \theta_0 \exp\{ikx \cos \theta_0\} \\ &\quad = \text{constant}\cdot\int_0^\infty \frac{[u(x_0)]}{x_0}\,dx_0 \int_{-i\infty}^{i\infty} \lambda^2 \cot \pi\lambda H_\lambda(x_0)J_\lambda(kx)\,d\lambda \\ &\quad = \text{constant}\cdot\int_{-i\infty}^{i\infty} \lambda[\lambda \cot \pi\lambda\bar{u}(\lambda)]J_\lambda(kx)\,d\lambda. \end{aligned} \tag{4.4}$$

Hence by (4.2) and (4.3),

$$\lambda \cot \pi\lambda\bar{u}(\lambda) \propto ik \sin \theta_0 \int_0^\infty \exp\{ik(x \cos \theta_0)\}H_\lambda(kx)\,dx.$$

The latter integral is obtained from the formulae (cf. [10]),

$$\int_0^\infty J_\mu(at)\begin{Bmatrix}\cos bt\\ \sin bt\end{Bmatrix} dt = \begin{Bmatrix}\cos\\ \sin\end{Bmatrix}(\mu \arcsin b/a), \qquad a > b, \qquad \Re e\mu > -1$$

We find

$$\lambda\bar{\mu}(\lambda) \propto \exp\{-i\lambda\pi/2\}\frac{\sin \lambda\theta_0}{\cos \pi\lambda} \tag{4.6}$$

or

$$[u(x_0)] \propto \int_{-i\infty}^{i\infty} \exp\{-i\lambda\pi/2\} \frac{\sin \lambda\theta_0}{\cos \lambda\pi} J_\lambda(kx)\, d\lambda. \tag{4.7}$$

The result is readily evaluated by residues.

(*5*) We now wish to make a few remarks with respect to the problem of the diffraction of a plane wave by a staggered array of semi-infinite planes, a problem solved by Carlson and Heins [2] by reducing it via Green's theorem to a Wiener-Hopf integral equation. The geometry is depicted in Figure 2.

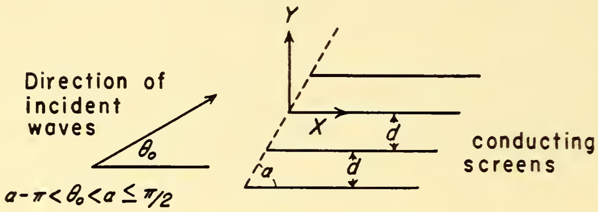


FIGURE 2

Separation of variables does not seem appropriate here at first sight, but in view of section (*2*), we may expect to find that a formulation in this way should succeed in a manner parallel to the Green's function integral equation. To see this we represent the perturbation due to each plate (in a Cartesian coordinate system whose origin is at the leading edge of that plate) in the manner of (3.1) as

$$u_m = \pm\frac{1}{2} \int_{-\infty}^{\infty} f_m(\beta) \exp\{i[\beta x_m + (k^2 - \beta^2)^{1/2} \mid y_m \mid]\}\, d\beta \tag{5.1}$$

with

$$x_m = x - md \cot \alpha, \qquad y_m = y - md.$$

We set

$$u = \exp\{ik(x \cos \theta_0 + y \sin \theta_0)\} + \sum_{m=-\infty}^{\infty} u_m\,. \tag{5.2}$$

Then, in order to avoid a discontinuity in u at $y = nd$ (off the plate), we require by (5.1) that $u_n = 0$ at $y_n = 0$ for x_n negative. Hence, $f_m(\beta)$ must be regular in a lower half-plane and of algebraic growth. We make a periodicity assumption on the current density, as done by Carlson and Heins, viz.,

$$u_m(x_m) = u_0(x) \exp\{ikm(d \cot \alpha \cos \theta + d \sin \theta)\}. \tag{5.3}$$

By (5.1)

$$f_m(\beta) = \int_{-\infty}^{\infty} u_m(x_m) \exp\{-i\beta x_m\}\, dx_m$$

$$= \int_0^{\infty} u_m(x_m) \exp\{-i\beta x_m\}\, dx_m$$

or

(5.4) $$f_m(\beta) = f_0(\beta) \exp\{ikm(d \cot \alpha \cos \theta + d \sin \theta)\}.$$

The condition $\partial u/\partial y = 0$ on metal, may be written

$$\left.\frac{\partial u}{\partial y}\right|_{y=y_m} = 0, \qquad \text{for} \qquad x_m > 0.$$

This yields,

(5.5) $$0 = \int_{-\infty}^{\infty} \left[\frac{-ik \sin \theta_0}{\alpha - k \cos \theta_0} + f_0(\beta)K(\beta)\right] \exp\{i\beta x_n\}\, d\beta = 0, \quad \text{for} \quad x_n > 0,$$

where $K(\beta)$ is obtained from (5.4) and (5.1) as

(5.6)
$$K(\beta) = i(k^2 - \beta^2)^{1/2} \sum_{p=-\infty}^{\infty} \exp\{ip(k\rho - \beta q) + i\,|\,p\,|\, d(k^2 - \beta^2)^{1/2}\}$$
$$= \frac{(k^2 - \beta^2)^{1/2} \sin d(k^2 - \beta^2)^{1/2}}{\cos d(k^2 - \beta^2)^{1/2} - \cos(k\rho - \beta q)}$$

with $q = d \cot \alpha$, $\rho = q \cos \theta + d \sin \theta$.

If we define $D(\beta)$ to be the integrand of (5.5), then $D(\beta)$ is to be regular in an upper-plane and of algebraic growth. The situation is then as in Carlson and Heins [2]. The use of the function theoretic techniques is necessary and suffices. It may be added that similar considerations apply to the case of a pair of semi-infinite plates (cf. [3]), with similar results.

(*6*) The examples hitherto discussed or mentioned deal with infinite boundaries and depend insofar as their Wiener-Hopf character is concerned, upon the use of Cartesian coordinates. This is the case in the work of Carrier, and essentially also in the treatment of the semi-infinite cylinder, by Schwinger and Levine [4].[2] In the present section, a finite obstacle will be considered. We have found that the semi-infinite plane constitutes a two part problem in Cartesians, but a one part problem in polars. The problem of a ribbon (under the conditions of section *1*), represented in cross-section by

$$y = 0, \qquad 0 < x < 1,$$

[2]The latter employ cylindrical coordinates r, θ, z, and the transform relates to the "Cartesian" z coordinate.

is, however, a *three part* problem in Cartesians, and the integral equation which results is a finite Wiener-Hopf equation. But, in polar coordinates, the problem is a two part problem. We sketch here the essential considerations in the case of Laplace's equation, in a very simple case which yet exhibits the essential features. We seek a function u such that it is regular together with its derivatives except on the ribbon. On the latter u may be discontinuous, but not $\partial u/\partial y$. On the ribbon $\partial u/\partial y$ is to be zero. More complicated solutions may be represented as integrals of this one with respect to x.

A solution by separation of variables is

$$u = \pm \int_{-i\infty}^{i\infty} r^{-\nu}[A(\nu) \cos \nu\theta + B(\nu) \sin \nu\theta] \, d\nu. \tag{6.1}$$

(An incident field may be represented in this way by means of the Mellin transform theorem, and included in more complex problems.) Continuity of $\partial u/\partial y$ at $\theta = 0$ and $\theta = 2\pi$, requires essentially that $B(\nu) = -A(\nu) \cot \pi\nu$. Continuity of u for $y = 0$, $x > 1$ leads to

$$\int_{-i\infty}^{i\infty} r^{-\nu} A(\nu) \, d\nu = 0, \qquad r > 1 \tag{6.2}$$

which in turn requires $A(\nu)$ regular in a right half-plane $\Re e\, \nu \geq 0$. The condition $\partial u/\partial y = 0$ on the ribbon leads to

$$\int_{-i\infty}^{i\infty} r^{-\nu} \nu B(\nu) \, d\nu = 0, \qquad r < 1 \tag{6.3}$$

so that $\nu B(\nu)$ is regular for $\Re e\, \nu \leq 0$, for reasons similar to those of section *2*. Thus we have

$$[\nu B(\nu)]_- = -A_+(\nu)\nu \frac{\cos \pi\nu}{\sin \pi\nu}. \tag{6.4}$$

Using the identities

$$\begin{aligned} \frac{\pi}{\sin \pi\nu} &= \Gamma(\nu)\Gamma(1-\nu) \\ \frac{\pi}{\cos \pi\nu} &= \Gamma\left(\frac{1}{2}+\nu\right)\Gamma\left(\frac{1}{2}-\nu\right), \end{aligned} \tag{6.5}$$

we find

$$[\nu B(\nu)]_- \frac{\Gamma(\frac{1}{2}-\nu)}{\Gamma(1-\nu)} = -A_+(\nu) \frac{\Gamma(1+\nu)}{\Gamma(\frac{1}{2}+\nu)}, \tag{6.6}$$

the left side being regular in the closed left half-plane and the right side in the closed right half. An entire function is thus defined, whose growth we study.

The gamma function quotients may be estimated by Stirling's theorem which leads to the formulae,

$$\frac{\Gamma(\frac{1}{2} - \nu)}{\Gamma(1 - \nu)} \sim (-\nu)^{-1/2},$$

$$\frac{\Gamma(1 + \nu)}{\Gamma(\frac{1}{2} + \nu)} \sim \nu^{+1/2}.$$

Hence, if in our case we let $\nu B(\nu)$ have a growth of the order $(-\nu)^{1/2}$, and $A(\nu) \sim \nu^{-1/2}$, the entire function is a constant and this determines $A(\nu)$ and $B(\nu)$. There is a relation between the growth of these functions and the behaviour of u and $\partial/\partial y$ near $r = 1$. We have

$$A(\nu) = \int_0^1 ur^{\nu-1}\, dr, \tag{6.7}$$

$$\nu B(\nu) = \int_1^\infty \frac{\partial u}{\partial y} r^\nu\, dr. \tag{6.8}$$

For large ν with positive real part, the crucial part of the integral in (6.7) comes from $r = 1$. Say u behaves like $(1 - r)^{-\beta}$ in that neighborhood. Then $A(\nu) \sim \int_0^1 r^{\nu-1}(1 - r)^{1-\beta-1}\, dr = \Gamma(\nu)\Gamma(1 - \beta)/(1 - \beta + \nu)$. A similar result holds with (6.8), after a change of variable $r = 1/\rho$.

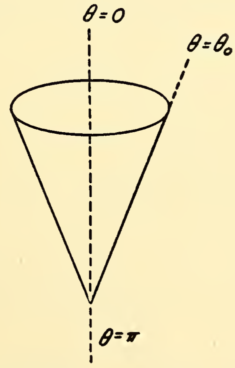


FIGURE 3

More complicated geometries of this type are readily constructed. Such problems can admittedly be solved by other means. The more interesting problem of the wave equation and this geometry is still unsolved, due to the complex nature of the normalization of the Bessel functions as functions of ν.

(7) Another problem which can be treated in this manner is the problem of a truncated cone. The geometry is indicated in Figure 3.

One may ask, for instance, for the electrostatic charge distribution on such an obstacle. We have then $\Delta u = 0$. Introducing the customary spherical coordinates and their product solutions, we write

$$u = \int_{\sigma - i\infty}^{\sigma + i\infty} r^{-\nu} A(\nu) P_{\nu-1}(+\cos\theta)\, d\nu, \qquad 0 < \theta < \theta_0, \tag{7.1}$$

$$u = \int_{\sigma - i\infty}^{\sigma + i\infty} r^{-\nu} B(\nu) P_{\nu-1}(-\cos\theta)\, d\nu, \qquad \theta_0 < \theta < \pi, \tag{7.2}$$

these representations being designed to avoid the singularities of the Legendre function. We require u continuous with its derivatives except on the conical cup, where $\partial u/\partial\theta$ is discontinuous. On the cup, $u = u_0$, a constant. (Problems involving oblique fields lead instead to sums of the form $u = \sum u_m \exp\{im\phi\}$, with functions $P_{\nu-1}^m$ (cos θ) occurring in u_m.) Continuity of u at $\theta = \theta_0$ implies

$$A(\nu)P_{\nu-1}(+\cos\theta_0) = B(\nu)P_{\nu-1}(-\cos\theta_0), \tag{7.3}$$

while continuity of $\partial u/\partial\theta$ for $\theta = \theta_0$ with $r > 1$ leads to

$$\begin{aligned} &\int_{\sigma - i\infty}^{\sigma + i\infty} r^{-\nu}\left\{A(\nu)\frac{dP_{\nu-1}}{d\theta}(+\cos\theta)\Big|_{\theta=\theta_0}\right. \\ &\qquad \left. - B(\nu)\frac{dP_{\nu-1}}{d\theta}(-\cos\theta)\Big|_{\theta=\theta_0}\right\} d\nu = 0, \qquad r > 1, \end{aligned} \tag{7.4}$$

or, defining $\phi_+(\nu)$ and utilizing (7.3) and the Wronskian of the pair of Legendre functions,

$$\begin{aligned} \phi_+(\nu) &\equiv A(\nu)P'_{\nu-1}(+\cos\theta_0) - B(\nu)P'_{\nu-1}(-\cos\theta_0) \\ &= \text{const.}\, B(\nu)\frac{\sin\nu\pi}{P_{\nu-1}(+\cos\theta_0)} = \text{const.}\, A(\nu)\frac{\sin\nu\pi}{P_{\nu-1}(-\cos\theta_0)}, \end{aligned} \tag{7.5}$$

Regular for $\Re e\nu \geq \sigma$,

Also the condition $u(r, \theta_0) = u_0$ for $0 < r < 1$ leads to the relation,

$$\int_{\sigma - i\infty}^{\sigma + i\infty} r^{-\nu}\left\{A(\nu)P_{\nu-1}(\cos\theta_0) - \frac{u_0}{\nu}\right\} d\nu = 0, \qquad r < 1 \tag{7.6}$$

when u_0 is expressed as a Mellin transform in the range of interest. The integrand must hence be regular in a left hand plane $\Re e\, \nu \leq \sigma$ and if we call it $g_-(\nu)$ we have, using (7.5),

$$\text{const.}\, \varphi_+(\nu)\frac{P_{\nu-1}(\cos\theta_0)P_{\nu-1}(-\cos\theta_0)}{\sin\pi\nu} - \frac{u_0}{\nu} = g_-(\nu). \tag{7.7}$$

(The analysis has been completed, and will be detailed elsewhere. We

content ourselves here with brief indications of its nature, together with the exposition of an interesting special case.) We introduce the subscripts $-\frac{1}{2} + \mu$ instead of $\nu - 1$; then $P_{1/2+\mu}(\cos\theta)$ is an entire function of μ, and its zeroes in the μ plane are real, and are ultimately in arithmetic progression for large positive μ. Also if μ is a zero so is $-\mu$. These facts permit the factorization of the coefficient of $\varphi_+(\nu)$ in (7.7) in the customary form $K_+(\nu)/K_-(\nu)$, as well as an estimate of the growths of these factors by comparison with the gamma function. The introduction of a suitable exponential factor into $K_+(\nu)$ and $K_-(\nu)$ results in algebraic growth. The analysis in general follows the more perspicuous lines of the special case $\theta_0 = \pi/2$, which represents a disc. For this special case, which we treat herewith, we have

$$P_{\nu-1}(0) = \frac{\pi^{1/2}}{\Gamma(1/2 + \nu/2)\Gamma(1 - \nu/2)};$$

expressing $\sin \pi\nu$ in terms of $\nu/2$ and using (6.6), (7.7) becomes

$$\varphi_+(\nu) \cdot \frac{\Gamma(\nu/2) \cdot \Gamma(1/2 - \nu/2)}{\Gamma(1/2 + \nu/2)\Gamma(1 - \nu/2)} - \frac{u_0}{\nu} = g_-(\nu). \tag{7.8}$$

This can be recast in the form

$$\begin{aligned} g_-(\nu)\,\frac{\Gamma(1 - \nu/2)}{\Gamma(1/2 - \nu/2)} + \frac{u_0}{\nu}\left\{\frac{\Gamma(1 - \nu/2)}{\Gamma(1/2 - \nu/2)} - \frac{\Gamma(1)}{\Gamma(1/2)}\right\} \\ = \frac{-u_0\Gamma(1)}{\nu\Gamma(1/2)} + \frac{\varphi_+(\nu)\Gamma(\nu/2)}{\Gamma(1/2 + \nu/2)}. \end{aligned} \tag{7.9}$$

If we now require in equation (7.1) $\sigma < 2$, then an entire function is defined. This amounts to an assumption on the behaviour of u at infinity. If $\phi_+(\nu) \sim \nu^\beta$ with $\beta < \frac{1}{2}$, and $g_-(\nu) \sim \nu^\alpha$ with $\alpha < -\frac{1}{2}$, the entire function is zero, by analysis like that of section *6*, the significance of such assumptions having been discussed there. $\phi_+(\nu)$, $A(\nu)$ and $B(\nu)$ are now determined. This result agrees with one obtained by Titchmarsh [11], who started out with the equations (obtainable from cylindrical coordinates),

$$\int_0^\infty f(u) J_0(\rho u)\, du = u_0, \qquad 0 < \rho < 1,$$

$$\int_0^\infty u f(u) J_0(\rho u)\, du = 0, \qquad \rho > 1,$$

but eventually employed Mellin transforms via a representation of $J_0(\rho u)$. His treatment of these equations has been improved by A. E. Heins [12].

A similar treatment of the disc problem for the wave equation case suggests itself, subject to the difficulties mentioned in section (*6*).

Bibliography

1. J. S. Schwinger, *Fourier transform solution of integral equations*, M.I.T. Radiation Laboratory Report.
2. J. F. Carlson and A. E. Heins, *The reflection of an electromagnetic plane wave by an infinite set of plates*, Part I, Quarterly of Applied Mathematics, Volume IV, 1946, pp. 313–330. Part II, Quarterly of Applied Mathematics, Volume V, 1947, pp. 82–89.
3. A. E. Heins, *The radiation and transmission properties of a pair of semi-infinite parallel plates*, Part I, Quarterly of Applied Mathematics, Volume VI, 1948, pp. 157–167. Part II, Quarterly of Applied Mathematics, Volume VI, 1948, pp. 215–220.
4. H. Levine and J. S. Schwinger, *On the radiation of sound from an unflanged circular pipe*, Physical Review, Volume 73, 1948, pp. 383–406.
5. G. Carrier, *A generalization of the Wiener-Hopf technique*, Quarterly of Applied Mathematics, Volume 7, 1949, p. 105.
6. E. T. Copson, *On an integral equation arising in the theory of diffraction*, Quarterly Journal of Mathematics, Oxford Series, Volume 17, 1946, pp. 19–34.
7. E. C. Titchmarsh, *Eigenfunction Expansions*, Oxford, London, 1946.
8. C. J. Bouwkamp, *A note on singularities occurring at sharp edges in EM diffraction theory*, Physica, Volume 12, October, 1946, pp, 467–74.
9. M. J. Kontorowich and N. N. Lebedev, *Uber eine methode zur lösung einigen probleme der beugung's theorie*, Journal of Physics, Moscow, Volume 1, 1939, pp. 229–241.
10. G. N. Watson, *Bessel Functions*, Macmillan, New York, 1945, p. 405, equations (4) and (5).
11. E. C. Titchmarsh, *Theory of Fourier Integrals*, Oxford, London, 1937, pp. 334–337.
12. A. E. Heins, *A note on a pair of dual integral equations*, Bulletin of the American Mathematical Society, Volume 56, No. 2, 1950, p. 172.
13. A. E. Heins, *Equation intégrales, Sur les couples d'equations intégrales*. Comptes Rendus de l'Académie des Sciences, Paris, Volume 230, 1950, pp. 1732–1734.

Asymptotic Solutions of a Differential Equation in the Theory of Microwave Propagation

By RUDOLPH E. LANGER
University of Wisconsin

The purpose of this paper is to show that asymptotic formulas for the solutions of a differential equation that is central to the theory of microwave propagation may be readily derived from results that are available in the mathematical literature. The motivation for this paper comes from one by C. L. Pekeris,[1] in which certain derivations of such formulas are made on a basis of power series methods. Although that approach is suggestive, it has its questionable aspects. Aside from the fact that it takes matters of convergence and rigor largely on faith, it seems fundamentally ill adapted, because of its essentially local character, for use in a problem which concerns an infinite range of the variable. It seems worth noting, because of this, that many needed formulas, dependably established, were already available in directly usable forms.

As to the organization of the paper, the formulation of the problem is briefly reviewed in section 1, and a criterion for distinction between configurations in which equation (1) must be differently dealt with is given. In section 2, the differential equation under conditions which apply to the "leaky" modes is discussed; especial attention is given to a method for deducing solutions that are explicit to more terms than the leading one. In section 3 the discussion is centered finally upon the differential equation under conditions which apply to the "transitional" modes.

In both sections 2 and 3 a qualitative comparison of the existing formulas with their analogues as obtained by power series methods is made. However, no actual quantitative check is included, since no such check has been undertaken.

1. *Introduction*

An analysis of the normal modes in microwave propagation in an atmosphere in which the index of refraction varies only with the height, is referable [P. 1108–

[1]C. L. Pekeris, *Asymptotic solutions for the normal modes in the theory of microwave propagation*, Journal of Applied Physics, Volume 17, 1946, pp. 1108–1124. This paper will be referred to by the designation P.

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Sciences and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directories of the Air Force Cambridge Research Laboratories.

10] in large measure to a determination of the forms of the solutions of a boundary problem in which the differential equation is of the form

$$U'' + k^2[\Lambda + y(h)]U = 0. \tag{1}$$

The variable h denotes the height, and is therefore positive and unbounded. Accents, such as those on U, indicate differentiations with respect to h. The coefficient $y(h)$ stands for the modified index of refraction, the variables having been changed from the natural ones to such as make the earth's surface flat. This function is thus also real and positive. The case to be especially considered is that in which it decreases over some initial range of h, as is indicated by the graph in Figure 1. A state of super-refraction, with consequent improved trans-

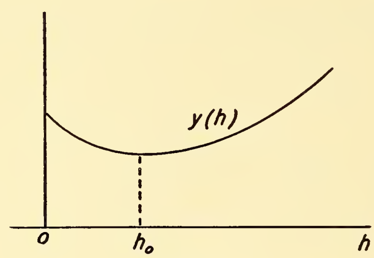


FIGURE 1

mission, then exists in the atmosphere's ground layer. The values k and Λ in the equation (1) are constants. Of these k is positive and fixed, and is assumed to be large. Λ, on the other hand, is to be regarded as an eigenvalue parameter, for characteristic values of which the differential equation has solutions that fulfill prescribed boundary conditions. In the paper P. these conditions are that U have the character of an outgoing wave when h is large, and that it vanish at $h = 0$.[2]

When a value (complex) of Λ is given, the equation

$$\Lambda + y(h) = 0, \tag{2}$$

clearly has no real root. However, if $y(h)$ is definable over an appropriate region of the complex plane, there may be a complex root h_1. If so, the equation (1) may be written in the form

$$U'' + k^2[y(h) - y(h_1)]U = 0, \tag{3}$$

with the result that it appears as a differential equation with a turning point at

[2]The present paper is not coextensive with the paper P. The latter goes into the determination of the eigenvalues Λ_m and the normalization of the corresponding solutions U_m, whereas the present paper is restricted to the asymptotic solutions of the differential equation alone.

h_1. A turning point is one at which the coefficient of the large parameter k^2 has a zero. The possibility of expressing equation (1) in the form (3) depends, of course, upon $y(h)$ being a function whose definition is extensible over a region which contains a root of the equation (2). We shall suppose that in the cases to be considered this is possible, specifically that $y(h)$ is thus extensible analytically over a region that is contiguous with the axis of reals. The forms which asymptotically solve the equation (1) relative to k depend upon the order of h_1 as a root of the equation (2), namely upon the order of the turning point.

It is the nature of an asymptotic formula relative to an unbounded parameter to be precise only in a limiting sense, and therefore to give only an approximation for any finite value of the parameter. In the case of equation (1), therefore, an asymptotic solution relative to k gives an approximation. The adequacy of this approximation for any particular purpose depends upon k being sufficiently large. Just what sufficient largeness may be is influenced by other factors, as is particularly clear in the case of equation (1), for k^2 enters this only in conjunction with the multiplier $[\Lambda + y(h)]$. Any degree of largeness of k^2 can therefore be vitiated by a corresponding smallness of this multiplier. Some normalization of this parameter is therefore called for. It may be made as follows:

If the turning point at h_1 is of the order n, so that $y^{[j]}(h_1) = 0$ for $1 \leqq j < n$, and $y^{[n]}(h_1) \neq 0$, the product $k^2[\Lambda + y(h)]$ may be written in the form

$$k^2 \frac{y^{[n]}(h_1)}{n!} \left[\frac{y(h) - y(h_1)}{\dfrac{y^{[n]}(h_1)}{n!}(h - h_1)^n} \right] (h - h_1)^n.$$

Here the factor $(h - h_1)^n$ is merely specific of the order of the turning point. It is, therefore, accounted for in the very design of the appropriate asymptotic solutions. The factor within brackets, which embodies the whole remaining dependence of the quantity upon h, has been constructed to have at h_1 the limiting value 1. It has thus been normalized to be of moderate magnitude near h_1. The constant $k^2 y^{[n]}(h_1)/n!$ is thus left to fill the role of the parameter. It is accordingly this, rather than k^2 alone, which must be sufficiently large if adequate approximations by the asymptotic formulas are to be assured.

These considerations are important in the analyses of the differential equation below. In section 2 the equation is dealt with when the turning point h_1 is of the first order. This point, however, is variable, due to its dependence upon Λ, whose eigenvalues are only subsequently to be determined. The condition for adequacy of the asymptotic solutions is in this case that $k^2 y'(h_1)$ be sufficiently large, and since the adequacy must obtain *uniformly* as to Λ it is clear that the condition must be *uniformly* fulfilled. This operates to bar h_1 from some neighborhood of the point h_0 of Figure 1, since at that point $y'(h)$ is zero. The formulas of section 2 are thus dependable only for values of Λ for

which h_1 does not lie within that neighborhood. For the excluded values, which are those that are associated with the transitional modes, the differential equation is then considered in section 3. The turning point is there located at the fixed position h_0. It is assumed that $y''(h_0) \neq 0$; more precisely that $k^2 y''(h_0)/2$ is sufficiently large.

2. *The Turning Point of the First Order*

If Λ is any value for which equation (2) has a simple root h_1, the differential equation (1) is expressible in the form (3) and thus has a turning point of the first order at h_1. The asymptotic solutions of such equations are known,[3] and are conveniently expressed in terms of the following variables:

$$\begin{aligned} \varphi(h) &= [\Lambda + y(h)]^{1/2}, \\ \Phi(h) &= \int_{h_1}^{h} \varphi(h)\, dh, \\ \Psi(h) &= \Phi^{1/6}(h)\varphi^{-1/2}(h), \\ \theta(h) &= \Psi''(h)\Psi^{-1}(h), \\ Q &= k\varphi(h), \\ u &= \int_{h_1}^{h} Q\, dh. \end{aligned} \tag{4}$$

The functions

$$z_j(h) = \Psi(h)u^{1/3}H_{1/3}^{(j)}(u), \qquad j = 1, 2, \tag{5}$$

in which the symbols $H_{1/3}^{(j)}$ stand for the Bessel functions (Hankel functions) usually so denoted,[4] are solutions of the differential equation

$$z'' + [k^2\varphi^2(h) - \theta(h)]z = 0. \tag{6}$$

Although the formula (4) for $\Psi(h)$ seems to assign a singularity to this function at h_1, it is found that this singularity is removable, and that $\Psi(h_1)$ is then different from zero. The function $\theta(h)$ is thus bounded in a region about h_1. Since the given differential equation (1) differs from (6) only to the extent of

---

[3] R. E. Langer, *On the asymptotic solutions of differential equations, with an application to the Bessel functions of large complex order*, Transactions of the American Mathematical Society, Volume 34, 1932, pp. 447–480. This paper will be referred to by the designation L_1.

[4] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge University Press, 1944, pp. 73–75.

this coefficient $\theta(h)$, it can be shown under suitable hypotheses that asymptotic solutions of (1) are obtainable from the known solutions (5) of equation (6).

We shall make this more precise, and in doing so shall adjust the formulations so as to permit Λ to appear not as a fixed but as a variable parameter. It is essential to do that if the formulas sought are to be usable in adjusting the solutions to boundary conditions. For, the eigenvalues of Λ for which the solutions fit given boundary conditions are usually unknown to begin with, and can be determined only reflexively from the forms which the solutions are found to have. For the discussion in this section $k^2 y'(h_1)$ must remain uniformly large. We shall therefore suppose that the range of Λ is a closed region of the complex plane for each point of which the equation (2) has a root h_1 at which $k^2 \mid y'(h_1) \mid$ exceeds some suitably large constant. As to this region, we shall suppose also that it includes no points of the axis of reals.[5] The function $y(h)$, being an index of refraction, is positive for positive h. We shall suppose that its definition is analytically extensible over some region of the complex plane that is contiguous with the axis of reals and includes all those values of h_1 which correspond to the admitted values of Λ.

Under these hypotheses the left-hand member of the equation (2) has a square root whose real part is positive and bounded from zero when h is positive. We shall understand that $\varphi(h)$ in (4) designates that root. We shall assume that the path of integration of the function $\Phi(h)$ called for in (4) extends from h_1 to some point of the axis of reals and then is taken along this axis. The real parts of $\Phi(h)$ and u are then increasing functions of h that become infinite with h. For large values of h, therefore, u is of large modulus and in the region

$$-\frac{\pi}{2} < \arg u < \frac{\pi}{2}.$$

We shall suppose, finally, that $y(h)$ is such that the integral

$$\int_0^\infty \left| \frac{\theta(h)}{\varphi(h)} \right| dh \tag{7}$$

is convergent. A wide class of cases fulfill this condition, in particular any one in which $y(h)$ has the character of either ch^ν or $ce^{\nu h}$ for all large positive h; c and ν being constants, with $\nu \geqq 0$.

The equation (1) then has [L_1, p. 459] a pair of solutions which maintain the forms

$$U_{0,1} = k^{-1/3}\varphi^{-1/2}(h)e^{iu}[1 + O(k^{-1})],$$

$$U_{0,2} = k^{-1/3}\varphi^{-1/2}(h)e^{-iu}[1 + O(k^{-1})],$$

[5]This includes the case of the leaky modes. The discussion could easily be made to include also real values of Λ. That would, however, require some additional hypotheses and a consideration of various cases in some of which the range of h would have to be restricted to exclude other turning points.

for large h, and are of the forms

$$U_{0,j} = z_j(h) + O(k^{-1}), \qquad j = 1, 2,$$

when h is such that $| u |$ is moderate or small. Upon taking U^* and U^{**} as suitable constant multiples of these, we therefore have the fact that there are solutions of equation (1) which have the forms

$$\begin{aligned} U^* &= (2/\pi)^{1/2} Q^{-1/2} e^{iu - 5\pi i/12}[1 + O(k^{-1})], \\ U^{**} &= (2/\pi)^{1/2} Q^{-1/2} e^{-iu + 5\pi i/12}[1 + O(k^{-1})], \end{aligned} \tag{8}$$

for large h, and are described by the formulas

$$\begin{aligned} U^* &= [u/Q]^{1/2} H_{1/3}^{(1)}(u) + O(k^{-4/3}), \\ U^{**} &= [u/Q]^{1/2} H_{1/3}^{(2)}(u) + O(k^{-4/3}), \end{aligned} \tag{9}$$

for moderate or small values of u. The solution U^{**} is clearly the one which has the form of an outgoing wave.

The formulas (8) and (9) are explicit only to the extent of their leading terms. If more precise forms are needed, such may be found by the following procedure.[6] In terms of the functions (5), let $\omega_1(h)$ and $\omega_2(h)$ be defined thus

$$\omega_j(h) = \left[1 + \frac{\alpha(h)}{k^2}\right] z_j(h) + \frac{\beta(h)}{k^2} z_j'(h), \qquad j = 1, 2, \tag{10}$$

with tentatively undetermined coefficients $\alpha(h)$ and $\beta(h)$. By differentiation and subsequent elimination of z_j'' through the use of equation (6), it is found that

$$\omega_j'(h) = \left[-\varphi^2\beta + \frac{\alpha' + \theta\beta}{k^2}\right] z_j + \left[1 + \frac{\alpha + \beta'}{k^2}\right] z_j' , \tag{11}$$

and a repetition of the process yields a corresponding formula for ω_j'' in terms of z_j and z_j'. Thus

$$\omega_j'' + k^2\varphi^2(h)\omega_j = \left[s_0 + \frac{s_1}{k^2}\right] z_j + \frac{s_2}{k^2} z_j' ,$$

[6] R. E. Langer, *The asymptotic solutions of ordinary linear differential equations of the second order, with special reference to a turning point*, Transactions of the American Mathematical Society, Volume 67, 1949, pp. 461–490. This paper will be referred to by the designation L_2. The method was also outlined in R. E. Langer, *On the connection formulas and the solutions of the wave equation*, Physical Review, Volume 51, 1937, pp. 669–676.

with

$$s_0 = -2\varphi^2\beta' - 2\varphi\varphi'\beta + \theta,$$

$$s_1 = \alpha'' + \theta\alpha + 2\theta\beta' + \theta'\beta,$$

$$s_2 = 2\alpha' + \beta'' + \theta\beta.$$

By the choice of $\beta(h)$, and then of $\alpha(h)$, we may make $S_0 = 0$ and $S_2 = 0$. The formulas for this are

$$\beta(h) = \frac{1}{2\varphi(h)} \int_{h_1}^{h} \frac{\theta(h)}{\varphi(h)}\, dh,$$

(12)

$$\alpha(h) = -\frac{1}{2}\left[\beta'(h) + \int_{h_1}^{h} \theta(h)\beta(h)\, dh\right],$$

and it is to be observed that, although $\beta(h)$ appears to be singular at h_1, this singularity is removable. With the determinations (12) we have, therefore,

$$\omega_j'' + k^2\varphi^2(h)\omega_j = (s_1/k^2)z_j, \qquad j = 1, 2. \tag{13}$$

Equations (10) and (11) are solvable for z_j in terms of ω_j and ω_j', and yield

$$z_j = D^{-1}\left\{\left[1 + \frac{\alpha + \beta'}{k^2}\right]\omega_j - \frac{\beta}{k^2}\omega_j'\right\},$$

with

$$D = 1 + \frac{2\alpha + \beta' + \varphi^2\beta^2}{k^2} + \frac{\alpha^2 + \alpha\beta' - \alpha'\beta - \theta\beta^2}{k^4}.$$

The result of substituting this evaluation into equation (13) is the relation

$$\omega'' + \frac{\beta s_1}{k^4 D}\omega' + \left\{k^2[\Lambda + y(h)] + \frac{s_1}{k^2 D}\left[1 + \frac{\alpha + \beta'}{k^2}\right]\right\}\omega = 0, \tag{14}$$

with ω_j in the place of ω. Since this follows for both ω_1 and ω_2, it has been found that the functions (10), (12), are solutions of the differential equation (14). This equation differs from the given equation (1) only by a coefficient of the order of k^{-4} for ω' and one of the order of k^{-2} for ω. From this it can be shown [L₂] that there exist solutions of (1) which are subject to the descriptions

$$U_j = \omega_j(h)[1 + O(k^{-3})], \qquad j = 1, 2,$$

when h is large, and

$$U_j = \omega_j(h) + O(k^{-3}),$$

when u is moderate or small. On multiplying these by $k^{-1/3}$ and using the abbreviations

$$Z_j(h) = [u/Q]^{1/2} H_{1/3}^{(j)}(u), \qquad j = 1, 2, \tag{15}$$

the conclusion may be stated as follows: There exist solutions of the equation (1) which maintain the forms

$$\begin{aligned} U^* &= \left\{\left[1 + \frac{\alpha(h)}{k^2}\right] Z_1(h) + \frac{\beta(h)}{k^2} Z_1'(h)\right\}[1 + O(k^{-3})], \\ U^{**} &= \left\{\left[1 + \frac{\alpha(h)}{k^2}\right] Z_2(h) + \frac{\beta(h)}{k^2} Z_2'(h)\right\}[1 + O(k^{-3})], \end{aligned} \tag{16}$$

when h is large, and are given by the formulas

$$\begin{aligned} U^* &= \left[1 + \frac{\alpha(h)}{k^2}\right] Z_1(h) + \frac{\beta(h)}{k^2} Z_1'(h) + O(k^{-10/3}), \\ U^{**} &= \left[1 + \frac{\alpha(h)}{k^2}\right] Z_2(h) + \frac{\beta(h)}{k^2} Z_2'(h) + O(k^{-10/3}) \end{aligned} \tag{17}$$

when u is moderate or small. These are the more explicit versions of the formulas (8) and (9).[7,8]

---

[7]This method may be used [L_2, 472–3] to derive formal solutions that are explicit to any prescribed powers of $1/k$.

[8]The formula which was derived by Pekeris (see footnote 1) as the counterpart of the second formula (17), is [P, 20]

$$U_m = \frac{u^{1/2}}{Q^{1/2}} H_{1/3}^{(2)}(u) + \frac{3Au^{5/6}}{2k^{4/3}Q^{1/2}} H_{-2/3}^{(2)}(u) - \frac{Bu^{3/2}}{2k^2Q^{1/2}} H_{4/3}^{(2)}(u) + Q^{-1/2}(h)O(k^{-8/3}),$$

with coefficients A and B which are constants expressible in terms of the values of $y(h)$ and its first four derivatives at h_1. This can be given a form much more similar to that of U^{**}. For (see footnote 4) it can be shown that

$$\begin{aligned} \frac{u^{5/6}}{Q^{1/2}} H_{-2/3}^{(2)}(u) &= \frac{-\Psi}{k^{2/3}} \{\Psi' Z_2 - \Psi Z_2'\}, \\ \frac{u^{3/2}}{Q^{1/2}} H_{4/3}^{(2)}(u) &= \left\{\frac{1}{3} + \frac{1}{2}\left(\frac{\Phi}{\varphi}\right)'\right\} Z_2 - \frac{\Phi}{\varphi} Z_2'. \end{aligned}$$

Thus the formula can be written alternatively

$$U_m = \left[1 + \frac{f_1(h)}{k^2}\right] Z_2(h) + \frac{f_2(h)}{k^2} Z_2'(h) + Q^{1/2}(h)O(k^{-8/3}),$$

with

$$f_1(h) = \frac{-3A}{2} \Psi\Psi' - B\left\{\frac{1}{6} + \frac{1}{4}\left(\frac{\Phi}{\varphi}\right)'\right\},$$

3. *The Turning Point of the Second Order*

When Λ is on a range of values for which the root h_1 of equation (2) lies too near the point h_0 of Figure 1, so that $k^2y'(h_1)$ is inadequately large, the analysis must be cast along different lines. With c as a constant subsequently to be determined, let

$$h = x + (c/k).$$

From Taylor's formula we have then

$$y(h) = y(x) + \frac{c}{k}\,y'(x) + \frac{c^2}{k^2}\,y''\left(x + \frac{c_1}{k}\right),$$

with c_1 some value between 0 and c. We may, therefore, write

$$k^2[\Lambda + y(h)] = k^2[y(x) - y(h_0)] + k[k\{\Lambda + y(h_0)\} + cy'(x)] + c^2y''\left\{x + \frac{c_1}{k}\right\},$$

and thus, on setting

$$\lambda = ki,$$

$$\text{(18)}\qquad \begin{aligned} X_0^2(x) &= y(x) - y(h_0),\\ X_1(x) &= i[k\{\Lambda + y(h_0)\} + cy'(x)],\\ X_2(x) &= -c^2y''\{x + (c_1/k)\}, \end{aligned}$$

we may write (1) in the form

$$\text{(19)}\qquad \frac{d^2U}{dx^2} - [\lambda^2X_0^2(x) + \lambda X_1(x) + X_2(x)]U = 0.$$

$$f_2(h) = \frac{3A}{2}\,\Psi^2 + \frac{B\Phi}{2\varphi}.$$

There is thus no qualitative conflict between the two results. No quantitative check upon them has been made.

The method by which the results of the present paper were derived is completely different from that of P. The latter is based upon the power series for $y(h)$ about the point h_1. Through operations upon tbis series, the coefficients of the differential equation for $Q^{1/2}\,U$ as a function of u are found in terms of power series in u. By retaining only the leading terms of these series an approximate differential equation is found. Even leaving aside the question of convergence of the various series which are thus brought into play, it will be clear that such a deduction bases itself entirely upon the character of $y(h)$ at the point h_1, even though the identification of the solution which is an outgoing wave must be made by reference to arbitrarily large values of h.

It appears therefore as a differential equation with a turning point of the second order at h_0, since at this point the coefficient $X_0^2(x)$ of the highest power of the large parameter λ has a zero of that order. The function $X_1(x)$, although it involves the constant k, is not large; for k appears in it only with the multiplier $\{\Lambda + y(h_0)\}$, which is evidently small since it is expressible as the integral of $y'(h)$ over the range from h_1 to h_0 where it is small.

A differential equation (19) is defined[9] to be in the normal form if its coefficients fulfill the relation

$$\Big[3X_0'X_1' - 2X_0''X_1\Big]_{h=h_0} = 0.$$

In the case at hand this may be assured by assigning an appropriate value to the constant c, namely by setting

$$c = \frac{2k\{\Lambda + y(h_0)\}y'''(h_0)}{q[y''(h_0)]^2}.$$

Under suitable hypotheses the forms of the solutions of the differential equation are then known. For use in expressing them we shall set

$$\sigma = \frac{-ki\{\Lambda + y(h_0)\}}{2[2y''(h_0)]^{1/2}},$$

$$(20)\qquad \begin{aligned} \eta(x) &= \frac{X_1(x)}{X_0(x)} + \frac{2\sigma X_0(x)}{\int_{h_0}^x X_0(x)\,dx}, \\ \varphi(x) &= 2X_0(x) + \frac{\eta(x)}{\lambda}, \\ \xi &= \lambda \int_{h_0}^x \varphi(x)\,dx. \end{aligned}$$

The singularity of $\eta(x)$ at h_0 is removable, because of the normality of the differential equation, and when it is removed $\eta(h_0) = 0$. If the function $y(h)$ is such that $X_0^2(x)$ is bounded from zero except in the immediate neighborhood of h_0, and if certain other conditions of a somewhat similar nature are fulfilled [L_3, 93, 101], the following is assured [L_3, 105, 106]: equation (1) has solutions which are of the forms

$$U_{0,1} = \lambda^{-1/4}\varphi^{-1/2}(x)\xi^{-\sigma}e^{(1/2)\xi}[1 + O(\lambda^{-1})],$$

$$U_{0,2} = \lambda^{-1/4}\varphi^{-1/2}(x)\xi^{\sigma}e^{-(1/2)\xi}[1 + O(\lambda^{-1})],$$

---

[9] R. E. Langer, *The asymptotic solutions of certain linear ordinary differential equations of the second order*, Transactions of the American Mathematical Society, Volume 36, 1934, pp. 90–106. This paper will be referred to by the designation L_3.

when h is large, and such that

$$U_{0,1} = \lambda^{-1/4}\varphi^{-1/2}(h)e^{-(\sigma+1/4)\pi i}\left\{\frac{2i\pi^{1/2}}{\Gamma(\frac{1}{4}+\sigma)} M_{\sigma,1/4}(\xi)\right.$$

$$\left. + \frac{\pi^{1/2}}{\Gamma(\frac{3}{4}+\sigma)} M_{\sigma,-1/4}(\xi)\right\} + O(\lambda^{-1}\log\lambda),$$

$$U_{0,2} = \lambda^{-1/4}\varphi^{-1/2}(h)\left\{\frac{-2\pi^{1/2}}{\Gamma(\frac{1}{4}-\sigma)} M_{\sigma,1/4}(\xi)\right.$$

$$\left. + \frac{\pi^{1/2}}{\Gamma(\frac{3}{4}-\sigma)} M_{\sigma,-1/4}(\xi)\right\} + O(\lambda^{-1}\log\lambda),$$

when $|\xi|$ is moderate or small. The symbols $M_{\sigma,\pm\frac{1}{4}}$ are used here to denote the confluent hypergeometric functions that are customarily so designated.[10]

With U^* and U^{**} taken as suitable constant multiples of the $U_{0,j}$, and with the variables Q, u and τ defined as

$$(21)\qquad \begin{aligned} Q &= k[y(x) - y(h_0)]^{1/2}, \\ u &= \int_{h_0}^{x} Q\,dx, \\ \tau &= \int_{h_0}^{x} \eta\,dx, \end{aligned}$$

the conclusion may be restated as follows: the differential equation (1) has solutions which for large values of h maintain the forms

$$(22)\qquad \begin{aligned} U^* &= e^{(\sigma+1/4)\pi i}\frac{(2iu+\tau)^{-\sigma}}{(Q-\eta i/2)^{1/2}}e^{iu+(1/2)\tau}[1+O(k^{-1})], \\ U^{**} &= \frac{(2iu+\tau)^{\sigma}}{(Q-\eta i/2)^{1/2}}e^{-iu-(1/2)\tau}[1+O(k^{-1})], \end{aligned}$$

and are given for moderate or small values of u by the formulas

$$U^* = \left(Q - \frac{\eta i}{2}\right)^{-1/2}\left\{\frac{2i\pi^{1/2}}{\Gamma(\frac{1}{4}+\sigma)} M_{\sigma,1/4}(2iu+\tau)\right.$$

$$\left. + \frac{\pi^{1/2}}{\Gamma(\frac{3}{4}+\sigma)} M_{\sigma,-1/4}(2iu+\tau)\right\} + O(k^{-5/4}\log k),$$

---

[10] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 3rd ed., Cambridge University Press, 1920, p. 337.

$$U^{**} = \left(Q - \frac{\eta i}{2}\right)^{-1/2} \left\{ \frac{-2\pi^{1/2}}{\Gamma(\frac{1}{4} - \sigma)} M_{\sigma,1/4}(2iu + \tau) \right.$$

(23)

$$\left. + \frac{\pi^{1/2}}{\Gamma(\frac{3}{4} - \sigma)} M_{\sigma,-1/4}(2iu + \tau) \right\} + O(k^{-5/4} \log k).$$

Of these solutions, U^{**} is evidently the one which has the form of an outgoing wave.[11]

---

[11]The paper P gives in the place of the second formula (23) the following one:

$$U_m = Q^{-1/2}(h) \left\{ \frac{2\pi^{1/2}}{\Gamma(\frac{1}{4} - \kappa)} M_{\kappa,1/4}(2iu(h)) + \frac{\pi^{1/2}}{\Gamma(\frac{3}{4} - \kappa)} M_{\kappa,-1/4}(2iu(h)) \right\},$$

with κ differing from σ, thus

$$\kappa = \sigma - \frac{i[7y'''(h_0) - 9y^{\text{iv}}(h_0)]}{288k[2y''(h_0)]^{3/2}}.$$

The most conspicuous discrepancy between this and the second formula (23) is in the sign of the term in $M_{\kappa,1/4}$. One suspects a typographical error in this. There are, of course, also a number of other differences, all of which however would approach zero if k were to become infinite.

Criteria for Discrete Spectra

By K. O. FRIEDRICHS
New York University

In discussing the nature of the spectra of differential operators I shall confine myself primarily to ordinary differential operators of the second order. Later on I shall indicate how this discussion can be extended to partial differential operators. In fact, one of the main advantages of the approach I shall employ is that it is not restricted to ordinary differential operators but can to a large extent be employed for partial differential operators.

The differential operator L we shall consider is of the second order, and self-adjoint; it acts on functions $\phi(x)$ which are defined in an interval $\mathcal{g}$ and subjected to various admissibility conditions. One may then ask for eigenfunctions

$$\phi = u(\lambda; x) \tag{1}$$

of this operator. These eigenfunctions are certain solutions of the eigenvalue equation

$$L\phi = \lambda\phi; \tag{2}$$

it is desired to expand arbitrary functions $\phi(x)$ in terms of these eigenfunctions. Since the differential operator L is of the second order, it should be possible to write the eigenvalue equation in the familiar form

$$\frac{d}{dx}\,p(x)\,\frac{d}{dx}\,\phi(x) - q(x)\phi(x) + \lambda r(x)\phi(x) = 0, \tag{3}$$

in which three given continuous functions $p(x)$, $q(x)$, and $r(x)$ appear. In order that equation (3) be equivalent to equation (2) we must define the operator L by

$$L = r^{-1}(x)\left[-\frac{d}{dx}\,p(x) + \frac{d}{dx}\,q(x)\right]. \tag{4}$$

The coefficients $r(x)$ and $p(x)$ are supposed to be positive in the interior

---

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Sciences and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directories of the Air Force Cambridge Research Laboratories.

of the interval $\mathcal{I}$. If the interval $\mathcal{I}$ is finite and if the functions $r(x)$ and $p(x)$ are continuous and positive at the endpoints and if also $q(x)$ is continuous there, the operator L is said to be regular or of the Sturm-Liouville type. It is well known that for such regular operators a complete sequence of eigenvalues λ and eigenfunctions $u(\lambda; x)$ exists provided that appropriate boundary conditions are imposed at the endpoints. The sequence is called complete if arbitrary functions $\phi(x)$ of a certain class admit an expansion of the form

$$\phi(x) = \sum_{\lambda} a_\lambda u(\lambda; x) \tag{5}$$

with appropriate coefficients a_λ.

We are here concerned with a singular operator L, i.e. with one for which not all regularity conditions mentioned above are satisfied. Thus, the operator L is singular, if the interval extends to infinity in one or both directions, or if one of the functions $r(x)$, $p(x)$, $r^{-1}(x)$, $p^{-1}(x)$, $q(x)$ does not approach a finite value at one endpoint at least.

It is well known that there are singular operators L which do not possess a complete sequence of eigenvalues and eigenfunctions in a proper sense. Nevertheless, for these singular operators there exists an analogue to the expansion of arbitrary functions with respect to eigenfunctions. In this modified expansion, integration occurs instead of summation, involving "improper" eigenfunctions, $v(\lambda; x)$, which depend continuously on the eigenvalue λ in a certain set S of values λ. Thus the expansion formula is of the form

$$\phi(x) = \int_S b(\lambda) v(\lambda; x)\, d\lambda + \sum_{\lambda} c_\lambda u(\lambda; x). \tag{6}$$

The set S in which the improper eigenfunctions $v(x; \lambda)$ are defined is called the continuous spectrum of the operator L; the values λ occurring in the sum form the point spectrum. The functions $u(\lambda; x)$ will also be called "proper" eigenfunctions.

While it is true that for certain singular operators L only an expansion of the type (6) is possible, there are nevertheless certain classes of singular operators for which an expansion of the type (5) with respect to proper eigenfunctions is possible in the same way as for regular operators. We then say that the operator L possesses a pure point spectrum; if the eigenvalues form a sequence tending to infinity we speak of a totally discrete spectrum. It is desirable to be able to determine from the nature of the coefficients p, q, and r whether or not the spectrum of L is totally discrete. The present discussion is primarily concerned with criteria which enable one to do this. More generally, we shall be concerned with criteria which will enable one to decide whether or not the part of the spectrum which lies below a certain value λ_* of λ is discrete, i.e., consists of a finite number of point eigenvalues, so that the continuous spectrum S, if there is any, is confined to values $\lambda \geq \lambda_*$.

Before formulating such criteria in detail I should like to mention a typical problem of wave propagation in which one is naturally led to ask whether or not the spectrum of an operator L is totally or partially discrete. Let us assume that the propagation of a field quantity $\chi = \chi(x, z, t)$ in a stratified medium is governed by the differential equation

$$\rho(x) \frac{\partial^2}{\partial t^2} \chi = \frac{\partial}{\partial x} p(x) \frac{\partial}{\partial x} \chi + \frac{\partial}{\partial z} r(x) \frac{\partial}{\partial z} \chi, \tag{7}$$

involving positive coefficients $p(x)$, $r(x)$ and $\rho(x)$ which depend on one coordinate, x, but not on the other, z. Let us ask whether or not there are waves of the form

$$\chi(x, z, t) = e^{ikt} e^{-i\mu z} \phi(x), \tag{8}$$

having a given frequency k and progressing in the z-direction, with the velocity k/μ. The frequency k being given, the wave number μ is to be determined. The amplitude $\phi(x)$ of such waves evidently satisfies the differential equation

$$\frac{d}{dx} p(x) \frac{d}{dx} \phi(x) - \mu^2 r(x) \phi(x) + k^2 \rho(x) \phi(x) = 0, \tag{9}$$

which agrees with equation (3) if one sets

$$q(x) = -k^2 \rho(x), \qquad \lambda = -\mu^2, \tag{10}$$

assuming the frequency k to be fixed. It is now particularly desirable to know whether or not there are unattenuated modes of propagation corresponding to proper eigenfunctions $\phi(x) = u(\lambda; x)$ associated with discrete negative eigenvalues $\lambda = -\mu^2$. In other words, one is interested in knowing whether or not the spectrum is discrete below the value $\lambda = 0$ and how many negative point eigenvalues there are. The criteria to be discussed will, in general, enable one to answer this question.

A complete theory of the expansion of an arbitrary function with respect to proper or improper eigenfunctions of a regular or singular ordinary differential operator L was given by Weyl in 1909 [1]. Treatments of this problem, more or less differing from that by Weyl, partly less general, partly extending or simplifying it, were given by Stone [2], Friedrichs [3], Titchmarsh [4], and Kodaira [5]. Sufficient conditions for the discreteness of the spectrum, of a more or less special character, were given by these authors in the course of their investigations.

Before describing such criteria, the eigenvalue problem of the operator L should be formulated in a more precise manner. The functions $\phi(x)$ should be quadratically integrable in the sense that

$$(\phi, \phi) = \int_{\mathfrak{s}} r(x) \mid \phi(x) \mid^2 dx < \infty. \tag{11}$$

The manifold of all such functions will be denoted by $\mathfrak{H}$. The operator L, given by (4), is not applicable on all functions in $\mathfrak{H}$, but only on those which are continuously differentiable and such that the function $p(x)\ d\phi/dx$ also has a derivative. We require that the function $L\phi(x)$ also belong to $\mathfrak{H}$. The manifold of all functions $\phi(x)$ for which this is the case will be denoted by $\mathfrak{F}^0$. In order that the operator become hypermaximal[1] so that an expansion with respect to eigenfunctions is defined in a unique way it may be necessary to impose further conditions, *boundary conditions*, on the functions $\phi(x)$ in $\mathfrak{F}^0$.

Weyl discovered the important fact that there are just two cases possible for each endpoint: either one of a certain linear manifold of boundary conditions must be imposed, or, no boundary conditions should be imposed at all. If at each endpoint one of the admissible boundary conditions is imposed in the first case or none in the second case the operator L will become hypermaximal. Weyl distinguished these two cases as limit circle and limit point cases, terms that were suggested by the particular approach that he employed. I shall not discuss here any details about this matter; I shall only mention that it is naturally important to be able to deduce from the nature of the coefficients p, q, r, which of the two cases arises. Various criteria about this alternative were given by the authors mentioned.

In the following we assume that at each endpoint one of the permissible boundary conditions has been imposed if necessary.[2] The manifold of functions $\phi(x)$ in $\mathfrak{F}^0$ satisfying these conditions will be denoted by $\mathfrak{F}$. The proper eigenfunctions should then belong to the space $\mathfrak{F}$. This requirement implies, in particular, that the integral (11) should be finite for them.

In case the spectrum of the operator L is totally discrete according to the definition given above, there exists a sequence of eigenvalues λ which increase and tend to infinity, $\lambda \uparrow \infty$. Every function $\phi(x)$ in $\mathfrak{H}$ possesses an expansion of the form

$$\phi(x) = \sum_{\lambda} c_\lambda u(\lambda; x).$$

For functions $\phi(x)$ in $\mathfrak{F}$, in addition, the expansion

$$L\phi(x) = \sum_{\lambda} \lambda c_\lambda u(\lambda; x) \tag{12}$$

holds. Here the summation refers to the sequence of eigenvalues and convergence is understood to hold in the mean. The coefficients c_λ are given in terms of the function $\phi(x)$ through simple integrals. The spaces $\mathfrak{H}$, or $\mathfrak{F}$, are

[1]Instead of this term, introduced by von Neumann, the term "self-adjoint" was employed by Stone. We have not used the latter term in order to avoid confusion with the property of "formal" adjointness of a differential operator.

[2]For a convenient way of determining the appropriate boundary conditions, see Rellich [6].

also the largest manifolds of functions for which respectively the first, or both, of these expansions hold.

For the present purpose it is suitable to adopt the following definition of partial discreteness of the spectrum: *We say the spectrum of the operator L is discrete below a value λ^* of λ if to every value $\lambda' < \lambda^*$ there exists at most a finite number of mutually orthogonal eigenfunctions $u_\lambda(x)$ associated with eigenvalues $\lambda \leq \lambda'$ such that for every function ϕ in $\mathfrak{F}$ which is orthogonal to the eigenfunctions $u_\lambda = u(\lambda; x)$,*

$$(u_\lambda, \phi) = 0, \tag{13}$$

the inequality

$$(\phi, L\phi) \geq \lambda_*(\phi, \phi) \tag{14}$$

holds. It then follows from the general theory that if the spectrum is discrete below every value λ_*, it is totally discrete.

After these preparations we may formulate various criteria.

To this end it is convenient to introduce a function $h(x)$ which is positive and of the form

$$h = \left| \int \frac{dx}{p(x)} + C \right| \tag{15}$$

in the neighborhood of each endpoint, $x = x_-$ or $x = x_+$, of $\mathcal{I}$. The constants $C = C_-$ and $C = C_+$ should be so chosen that at each endpoint h is either zero or infinite. In other words, the positive function $h(x)$ should be so chosen that $|dh/dx| = 1/p$ near the end points and that it becomes either zero or infinite at the endpoints.

We further introduce the quantity

$$Z(x) = \frac{1}{r(x)} \left[q(x) + \frac{1}{4p(x)h^2(x)} \right]. \tag{16}$$

We maintain that the discrete character of the spectrum depends on the behavior of this quantity $Z(x)$ at the endpoints $x = x_-$ and $x = x_+$. The first criterion is

I. *The spectrum of the operator L is totally discrete if*

$$Z(x) \to \infty \quad \text{as} \quad x \to x_- \quad \text{and} \quad x \to x_+ . \tag{17}$$

Certainly, this test is easily applied. Let us, in particular, take the case that $r(x) = p(x) = 1$, $x_- = 0$, $x_+ = \infty$. Then $h(x) = x$, and the criterion reduces to

$$q(x) + \frac{1}{4x^2} \to \infty \quad \text{as} \quad x \to 0 \quad \text{and} \quad x \to \infty. \tag{17'}$$

This condition is certainly satisfied if $q(x) \to \infty$ as $x \to \infty$ while $q(x)$ remains bounded at $x = 0$. That the spectrum is discrete under this condition is a very familiar fact, which plays a considerable role in spectral problems of quantum theory. It is to be noted that condition (17′) admits also functions $q(x)$ which are negative infinite as $x \to 0$. A singularity of $q(x)$ at $x = 0$ like that of a Coulomb potential, $q(x) = -c/x$, for example, does not spoil the totally discrete character of the spectrum.

It may be mentioned that in principle it would be sufficient to formulate criteria under the condition $p(x) = r(x) = 1$, since this condition can be satisfied by introducing new dependent and independent variables instead of ϕ and x provided $p(x)$ and $r(x)$ possess second derivatives. However, when expressed in terms of the original coefficients p, q, and r, condition (17′) becomes very complicated and, therefore, it seems preferable to use the rather simple condition (16).

Relation (17) is not a necessary condition for totally discrete character, but it is almost a necessary condition, as seen from the following two criteria.

II. *The spectrum of the operator L is discrete below the value $\lambda = \lambda_*$ if*

$$\liminf Z(x) \geq \lambda_* \tag{18}$$

with reference to $x \to x_-$ and $x \to x_+$.

III. *The spectrum of the operator L is not discrete below λ_*'' , if*

$$\limsup Z(x) < \lambda_* \tag{19}$$

with reference to either $x \to x_-$ or $x \to x_+$, provided that $Z(x)$ is bounded below.

Consequently, if $Z(x)$ approaches a definite limit λ_{**} at one endpoint and has an inferior limit $\geq \lambda_{**}$ at the other endpoint the spectrum is discrete below λ_{**} but not below any value $\lambda_* > \lambda_{**}$. In other words, a non-discrete spectrum begins at $\lambda = \lambda_{**}$.

The conditions of these three criteria imply that the quantity $Z(x)$ is bounded below. As a consequence, as can be shown, also the spectrum of the operator L is bounded below. If, however, $Z(x)$ approaches $-\infty$ at one endpoint, the spectrum also extends to $-\infty$. This spectrum may be a continuous spectrum, but strangely enough, the spectrum may even be discrete. We then speak of a discrete spectrum unbounded below. More specifically, we say that *the spectrum of the operator L is discrete below λ_* , unbounded below, if there is a sequence of eigenvalues $\lambda < \lambda_*$ which decreases and approaches ∞ such that inequality* (14) *holds for all functions ϕ in $\mathfrak{F}$ which are orthogonal to all corresponding eigenfunctions, i.e. which satisfy condition* (13). If this is the case for every value λ_* , an increasing and decreasing sequence of eigenvalues exists such that the expansions (5) and (12) hold; we then say the spectrum is totally discrete unbounded below. We now state, slightly generalizing a criterion due to Rellich.

IV. *The spectrum of the operator L is discrete below* λ_*, *unbounded below, if* $\liminf Z(x) \geq \lambda_*$ *at one endpoint, e.g.* $x = x_-$, *while for the other endpoint,* $x = x_+$, *the relation*

$$Z(x) \to -\infty \qquad \text{as} \qquad x \to x_+ \tag{20}$$

holds in such a way that

$$\Omega = \int_{x_0}^{x_+} \frac{[r(x)]^{1/2}\,dx}{[-p(x)Z(x)]^{1/2}} < \infty, \tag{21}$$

x_0 *being any number such that* $Z(x) < 0$ *for* $x = x_0$. Of course, the roles of x_+ and x_- could be interchanged. (The validity of this criterion has so far been proved only if the functions p, q, r satisfy additional smoothness conditions; for such additional conditions in case $p = r = 1$, $x_+ = \infty$, see [4].)

In case $p(x) = r(x) = 1$, $q(x)$ continuous and negative, $0 < x < \infty$, and $x_+ = \infty$, conditions (20) and (21) may be replaced by

$$q(x) \to -\infty \qquad \text{as} \qquad x \to \infty \tag{20$'$}$$

and

$$\int_0^\infty \frac{dx}{[-q(x)]^{1/2}} < \infty. \tag{21$'$}$$

If, in addition, the function $q(x)$ is defined continuously at $x = 0$, the spectrum is totally discrete unbounded below.[3] If condition (20′) is satisfied but not (21′) the spectrum is continuous; this interesting statement is due to Titchmarsh [4].

As an example of the application of criterion IV we may take the case of spherical waves characterized, with reference to equation (9), by

$$p(x) = x^2, \qquad r(x) = 1, \qquad \rho(x) = x^2, \qquad x_- = 0, \qquad x_+ = \infty,$$

so that $q(x) = -k^2x^2$, $k \neq 0$, by (10) and

$$h(x) = x^{-1}, \qquad Z(x) = -k^2x^2 + \tfrac{1}{4}.$$

Hence

$$Z(x) \to \tfrac{1}{4} \qquad \text{as} \qquad x \to 0,$$

$$Z(x) \to -\infty \qquad \text{as} \qquad x \to \infty,$$

[3]This remarkable fact was mentioned to me by Rellich. *Addition in proofs:* It is also stated in a recent paper by Sears and Titchmarsh [9] correcting erroneous statements made in [4].

and

$$\Omega = \int_{k^{-1}}^{\infty} \cdot \frac{dx}{(k^2x^4 - x^2/4)^{1/2}} < \infty .$$

Consequently, the spectrum is discrete below $\lambda = 1/4$, unbounded below. The nondiscrete character of the spectrum above $\lambda = 1/4$ is due to the behavior at the origin.

Various different criteria have been formulated in the literature; some of them give more information in the case that the inferior and superior limits of the function $Z(x)$ at one endpoint differ from each other. I do not intend to mention here those fine points of difference. It is also possible to prove the validity of these criteria in many different ways. I shall indicate only one approach which has the advantage that it does not rely on the fact that the differential operator is an ordinary differential operator and which, therefore, to a large extent can be used for partial differential operators.

This approach is based on the fact that hypermaximal operators with a totally discrete spectrum can be characterized by an abstract property, which in most concrete cases is easily verified. This important property was discovered by Hilbert [7] in 1906 for the case of bounded operators, such as integral operators. Later on, Weyl [8] extended this property to operators with a partially discrete spectrum and also made use of it for differential operators, which are not bounded, after having transformed them into bounded integral operators with the aid of a Green's function.

It is, however, possible to utilize the ideas of Hilbert and Weyl for differential operators directly. In carrying this out one obtains at the same time upper and lower estimates for the number of discrete eigenvalues below an arbitrarily chosen value. On the other hand, this direct approach seems to be restricted to the case in which the quantity $Z(x)$, and hence the spectrum, is bounded below.

For simplicity we restrict ourselves to the case in which a particular boundary condition is imposed.

We consider the quadratic forms

$$(\phi, \phi) = \int_{x_-}^{x_+} r(x)\phi^2(x)\,dx,$$

and

$$(\phi G \phi) = \int_{x_-}^{x_+} \left[p(x)\left(\frac{d}{dx}\phi(x)\right)^2 + q(x)\phi^2(x) \right] dx, \tag{22}$$

further the space $\mathfrak{H}$ of all functions $\phi(x)$ for which (ϕ, ϕ) is finite and the space $\mathfrak{G}$ of all differentiable functions $\phi(x)$ for which $(\phi G\phi)$ is finite and which tend

to zero as x approaches an endpoint at which the function $h(x)$ vanishes, see (15). Then we state

A. *Suppose there are n functions* $\phi^{(1)}(x), \cdots, \phi^{(n)}(x)$ in $\mathfrak{H}$ *such that the inequality*

$$(\phi G \phi) + \sum_{\nu=1}^{n} (\phi, \phi^{(\nu)})^2 \geq \lambda'(\phi, \phi) \tag{23}$$

holds for all functions $\phi(x)$ *in* $\mathfrak{G}$. *Then the spectrum of the operator* L *is discrete below the value* λ' *and there are at most* n *eigenvalues of* L *below* λ'.

In case the inferior limit of the quantity $Z(x)$ at the endpoints is greater than λ', it is not very difficult to construct a finite number of functions $\phi^{(\nu)}(x)$ such that inequality (23) holds. In this way, the validity of criterion II and hence of I, can be established. At the same time an upper estimate, n, of the number of eigenvalues $\lambda < \lambda'$, is found.

The proof of criterion III for non-discreteness of the spectrum can be given in connection with the following fact:

B. *Suppose there are m linearly independent functions* $\psi^{(1)}(x), \cdots, \psi^{(m)}(x)$, *in* $\mathfrak{G}$ *such that for every linear combination*

$$\phi(x) = \sum_{\mu=1}^{m} c_\mu \psi^{(\mu)}(x) \tag{24}$$

which does not vanish identically,

$$\phi(x) \not\equiv 0, \tag{25}$$

the inequality

$$(\phi G \phi) < \lambda_*(\phi, \phi) \tag{26}$$

holds. Then either there are at least m point eigenvalues below λ_*, *or the spectrum is not discrete below* λ_*.

Under the conditions of criterion III it is not difficult to prove that there are arbitrarily many eigenvalues below a certain value $\lambda' < \lambda_*$. This fact implies that the spectrum is not discrete below λ_*. Thus the validity of criterion III can be established.

Statement B evidently enables one to obtain a lower estimate, m, of the number of eigenvalues below the value λ_*. We illustrate this fact with the rather trivial example in which $p(x) = r(x) = 1$ and $q = -\kappa^2$ for $0 \leq x < a$, while $q = 0$ for $a < x < \infty$. We simply take the functions $\psi^{(\mu)}(x) = \sin(\mu\pi x/a)$, $x \leq a$, $=0$, $x \geq a$, for $\nu = 1, \cdots, m$. For any linear combination $\phi(x) = \sum_{\mu=1}^{m} c_\mu \psi^{(\mu)}(x)$ of these functions we evidently have

$$(\phi, \phi) = \sum_{\mu=1}^{m} c_\mu^2 \frac{a}{2},$$

$$(\phi G \phi) = \sum_{\mu=1}^{m} \left(\frac{\nu^2\pi^2}{a^2} - \kappa^2\right) c_\mu^2 \frac{a}{2},$$

and hence

$$(\phi G \phi) \leq \left(\frac{m^2\pi^2}{a^2} - \kappa^2\right)(\phi, \phi).$$

From statement B we now infer that there are at least m negative eigenvalues if

$$m < \kappa a/\pi.$$

If one intends to employ the criteria A and B in concrete cases, one must find appropriate functions $\phi^{(\nu)}(x)$ and $\psi^{(\mu)}(x)$. To do this right requires skill and experience since there do not seem to exist hard and fast rules that tell one how to find such functions, although a few results of experience could be formulated.

These two criteria A and B have various ramifications. I only mention, as P. Lax has noticed, that the properties A and B are respectively equivalent with the converse properties $\overline{B}$ and $\overline{A}$, which can be put to good use. Also there is a close connection between properties A and B in Courant's maximum minimum property and a corresponding minimum maximum property.

It is rather evident that the criteria A and B can also be formulated with reference to partial differential operators; all that is needed is to define analogues of the forms (11, 22). As a consequence, the criteria I to IV can be carried over to the case of partial differential operators provided the functions p, q, r behave uniformly at each component of the boundary. Certainly it should be possible to relax even this condition. This extension to partial differential operators has, however, been carried out only in special cases.

Bibliography

1. Weyl, H., *Über gewöhnliche Differentialgleichungen mit singulären Stellen,* Göttinger Nachrichten von der Gesellschaft der Wissenschaften, 1909, pp. 37–64. *Über gewöhnliche Differentialgleichungen mit Singularitäten,* Mathematische Annalen, Volume 68, 1910, pp. 373–392. *Ramifications, old and new, of the eigenvalue problem,* Bulletin of the American Mathematical Society, Volume 56, 1950, pp. 115–139.
2. Stone, M. H., *Linear transformations in Hilbert space,* American Mathematical Society, Colloquium Publication, New York, 1932, Ch. X, p. 484, Theorem 10.19.
3. Friedrichs, K. O., *Über die ausgezeichnete Randbedingung in der Spektraltheorie der halbbeschränkten gewöhnlichen Differentialoperatoren zweiter Ordnung,* Mathematische Annalen, Volume 112, 1935, pp. 1–23. *On differential operators in Hilbert spaces,* American Journal of Mathematics, Volume 61, 1939, pp. 523–544, and earlier papers

quoted there. *Criteria for the discrete character of the spectra of ordinary differential operators*, Courant Anniversary Volume, Interscience, New York, 1948, pp. 145–160.

4. Titchmarsh, E. C., *Eigenfunction Expansions Associated with Second-Order Differential Equations*, Oxford, 1946, Ch. V, pp. 97–117.
5. Kodaira, K., *The eigenvalue problem for ordinary differential equations of the second order and Heisenberg's theory of S-matrices*, American Journal of Mathematics, Volume 71, 1949, pp. 921–945.
6. Rellich, F., *Die zulässigen Randbedingungen bei den singulären Eigenwertproblemen der mathematischen Physik*, Mathematische Zeitschrift, Volume 49, 1944, pp. 702–723.
7. Hilbert, D., *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, Nachrichten von der K. Gesellschaften der Wissenschaften zu Göttingen, vierte Mitteilung, 1906, pp. 157–227. See also the monograph of 1924, p. 147.
8. Weyl, H., *Über beschränkte quadratische Formen, deren Differenz vollstetig ist*, Rendiconti del Circolo Matematico di Palermo, Adunanza del 13, December, 1908, pp. 1–20.
9. Sears, D. B., and E. C. Titchmarsh, *Some eigenfunction formulae*, The Quarterly Journal of Mathematics, Oxford, Second Series, Volume 1, 1950, pp. 165–175.

Extension of Weyl's Integral for Harmonic Spherical Waves to Arbitrary Wave Shapes

By H. PORITSKY
General Engineering and Consulting Laboratory,
General Electric Company

1. *Introduction*

In his paper of 1919 on the propagation of radio waves generated by a dipole near a flat earth, H. Weyl utilized to advantage the representation of a sinusoidal spherical wave in terms of sinusoidal plane waves. With the point

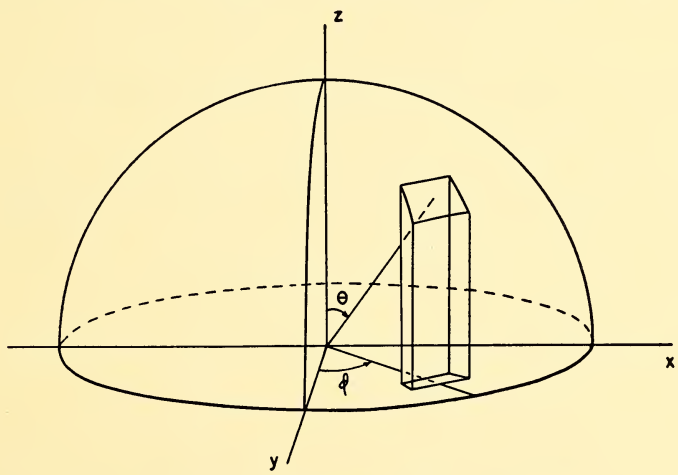


FIGURE 1

source of the spherical wave at the origin, this representation for $z > 0$ is given by the double integral

$$(1.1) \qquad \frac{e^{-ik_0R}}{-ik_0R} = \frac{1}{2\pi}\int \exp\{-ik_0(\alpha x + \beta y + \gamma z)\}\, d\omega, \qquad z > 0,$$

where (see Figure 1) the integration is carried out over the half sphere $\gamma > 0$ of the unit sphere

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

$$\alpha^2 + \beta^2 + \gamma^2 = 1, \tag{1.2}$$

as well as over the complex portion of the unit sphere (1.2) corresponding to θ varying from 0 to $\pi/2 + i\infty$. This means that if α, β, γ, $d\omega$ be expressed in spherical polar coordinates θ, φ:

$$\begin{aligned} \alpha &= \cos\varphi \sin\theta, \\ \beta &= \sin\varphi \sin\theta, \\ \gamma &= \cos\theta, \\ d\omega &= \sin\theta \, d\theta \, d\varphi, \end{aligned} \tag{1.3}$$

then the φ limits in (1.1) are from 0 to 2π, while the θ limits are not from 0 to $\pi/2$, but are given by

$$\left[\int_0^{\pi/2} + \int_{\pi/2}^{\pi/2+i\infty}\right] d\theta \tag{1.4}$$

as shown in Figure 2 in the complex θ-plane. These complex directions naturally do not show up in Figure 1.

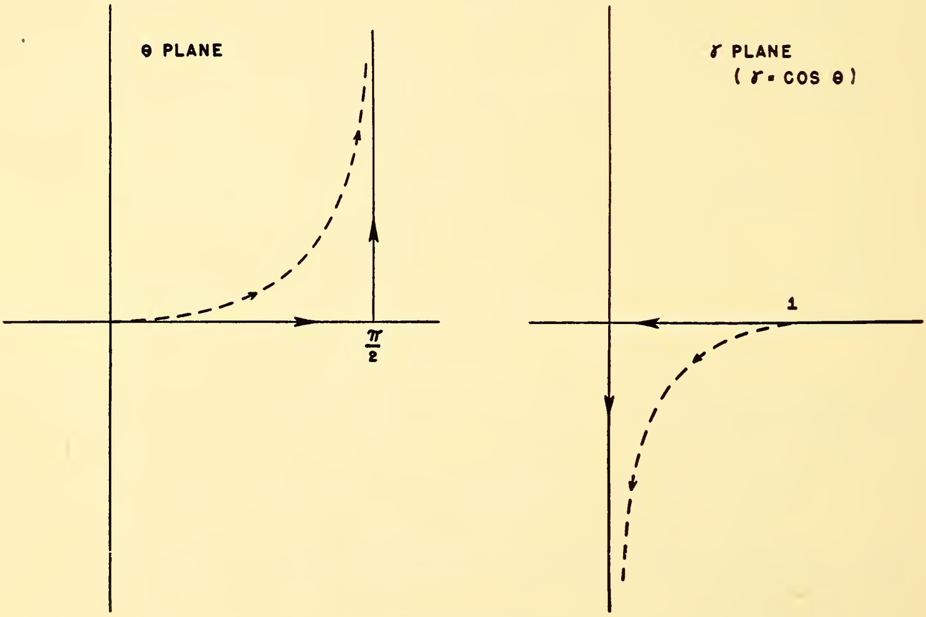


FIGURE 2

The corresponding values of γ vary from 1 to 0, then to $-i\infty$, as shown on Figure 2. Since the path of integration of an analytic function may be distorted

in the complex plane, the rectangular path may be changed into a path such as shown on Figure 2 by the broken line, from 0 to $\pi/2 + i\infty$.

By reflecting and refracting at $z = 0$ each plane wave component of a similar representation in $z < h$ of a point source placed at $(0, 0, h)$, Weyl obtained a similar double integral representation for the field of the dipole both above and in the ground.

In evaluating the integrals in question, Weyl makes very effective use of changes of variables along the sphere (1.2) in spite of the partly real, partly complex range of integration. Thus, in proving (1.1) for any point P: (x, y, z), he introduces spherical coordinates $R = (x^2 + y^2 + z^2)^{1/2}$, η, ψ with the line from the origin to the point P in question as polar axis, replaces the right-hand member of (1.1) by

$$(1.5) \qquad \frac{1}{2\pi}\int_0^{2\pi} d\psi \int_0^{\pi/2+i\infty} \exp\{-ik_0R\cos\eta\}\sin\eta\, d\eta,$$

and introduces $\tau = ik_0R\cos\eta$. The values of η, $\cos\eta$, τ over the path of integration are shown in Figure 3. The integral (1.5) becomes

$$(1.6) \qquad -\frac{1}{ik_0R}\int_{ik_0R}^{\infty} e^{-\tau}\, d\tau = \frac{1}{-ik_0R}\, e^{-\tau}\Big]_{+\infty}^{ik_0R} = \frac{e^{-ik_0R}}{-ik_0R}.$$

The shifting of variables to η, ψ is perhaps not as puzzling as the new choice of limits. This corresponds to the freedom of moving paths of inte-

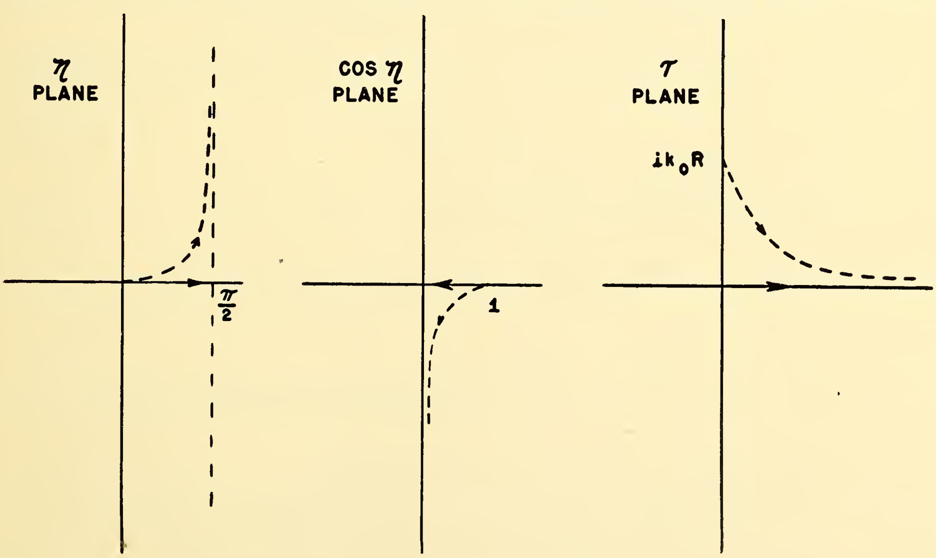


FIGURE 3

gration of a single variable in the complex plane; here the two-dimensional area of integration is distorted on the complex sphere (1.2), but without the benefit of a visual representation which is available for a single complex variable.

However, since a spherical wave possesses spherical symmetry to start with, a shift of axes from that of Figure 1 to a similar set with OP as polar axis, is possible, and an equation similar to (1.1) may be set up for coordinate axes with OP as the z-axis, leading to (1.5) directly, rather than by transforming (1.1) from θ, φ to η, ψ.

In the following, we consider an extension of the Weyl integral given by

$$\frac{F(ct - R)}{R} = -\frac{1}{2\pi} \int F'[ct - (\alpha x + \beta y + \gamma z)]\, d\omega. \tag{1.7}$$

The integral (1.7) represents a divergent spherical wave of arbitrary wave shape but with the inverse radius variation of amplitude, $F(ct - R)/R$, as a superposition of plane waves whose wave shape is F', the derivative of F, and of uniform amplitude per element of solid angle $d\omega$ presumably of the same half-sphere as in equation (1.1) and Figure 1, including both real and complex directions.

Since F is defined originally as an arbitrary real function of its real argument, whereas in (1.7) F' must range over complex values of its argument, the extension of F and F' to complex values must be first considered before (1.7) acquires a definite meaning.

2. *Derivation of (1.7) by Means of Fourier Integral Time Resolution, and by Assuming Analyticity of F*

One way of extending the Weyl integral to arbitrary spherical wave solutions

$$F(ct - R)/R \tag{2.1}$$

is by applying the Fourier integral in time.

Recalling the exponential time factor

$$e^{i\omega t}, \qquad k_0 = \omega/c, \tag{2.2}$$

write (1.1) in the form

$$\frac{\exp\{ik_0(ct - R)\}}{R} = -\frac{ik_0}{2\pi} \int \exp\{ik_0[ct - (\alpha x + \beta y + \gamma z)]\}\, d\omega. \tag{2.3}$$

Next resolve F in $F(ct - R)$ into a Fourier integral

$$F(ct - R) = \int_{-\infty}^{+\infty} A(k_0) \exp\{ik_0(ct - R)\}\, dk_0\,. \tag{2.4}$$

Utilizing (2.3) we obtain

$$F(ct - R) = \frac{R}{2\pi} \int_{-\infty}^{+\infty} -ik_0 A(k_0)\, dk_0 \int \exp\{ik_0[ct - (\alpha x + \beta y + \gamma z)]\}\, d\omega. \tag{2.5}$$

Now it will be noted that differentiation of (2.4) with respect to R yields

$$F'(ct - R) = \int_{-\infty}^{+\infty} ik_0 A(k_0) \exp \{ik_0(ct - R)\}\, dk_0 \,. \tag{2.6}$$

Hence, carrying out the k_0-integration in (2.5) first, one presumably arrives at (1.7).

In the derivation of (1.7) just given the Fourier integration over k_0 is carried out for real as well as complex values of argument $[ct - (\alpha x + \beta y + \gamma z)]$, due to the complex portion of the sphere of integration, as explained in Figure 2 and Section 1. Hence F', initially defined for real values, must be extended to complex values of its argument as an analytic function of its argument before equation (1.7) becomes unambiguous. This point will be examined in detail in Section 3. Moreover, it turns out that the integrals for negative k diverge.

If by accident F can be extended to complex arguments ζ as an analytic function $F(\zeta)$ in the half-plane $I(\zeta) \geqq 0$, then a *direct* proof of equation (1.7) may be given, based on change of variables from θ, φ to η, ψ utilized by Weyl in his proof of equation (1.1). With this change of variables, the right-hand member of (1.7) becomes

$$\begin{aligned} &-\frac{1}{2\pi} \int_0^{\pi/2 + i\infty} F'(ct - R \cos \eta) \sin \eta \, d\eta \int_0^{2\pi} d\psi \\ &= \int_1^{i\infty} F'(ct - R \cos \eta)\, d(\cos \eta) \end{aligned} \tag{2.7}$$

and introducing

$$\tau = -R \cos \eta \tag{2.8}$$

$$\frac{1}{R} \int_{i\infty}^{-R} F'(ct + \tau)\, d\tau = \frac{F(ct - R) - F(i\infty)}{R}. \tag{2.9}$$

In order that the right-hand member of (2.9) reduce to the left-hand member of (1.7), it is necessary that

$$F(i\infty) = 0, \tag{2.10}$$

and it will be sufficient, if more generally

$$F(\infty) = 0. \tag{2.11}$$

Indeed, since the derivative F' determines F only to within an additive constant, it is clear that equation (1.7) cannot possibly hold without some restriction on F, and the condition (2.11) will be used whenever possible.

Upon closer examination, both proofs of (1.7) given above will be found insufficient. In the first place, the proof by means of the Fourier time resolution involves components with negative k_0, and for such k_0 the exponentials factor $\exp \{-ik_0\gamma z\}$ in (1.1) leads to divergent integrals. Similarly, the direct proof

based on the analyticity of F in the upper half-plane of its argument and the assumption (2.11), does not apply, in general, to cases when F is given as an arbitrary real function of its real argument.

It is shown in Section 3 that while in general it is impossible to extend an arbitrary function $F(\xi)$, defined for real values of ξ, to complex values $\zeta = \xi + i\eta$, as an analytic function of ζ over the whole ζ-plane, it is always possible to break up $F(\xi)$ into a sum of two functions

$$F(\xi) = \tfrac{1}{2}[F_1(\xi) + F_2(\xi)], \tag{2.12}$$

such that $F_1(\xi)$ can be extended as an analytic function $F_1(\zeta)$ in $\eta > 0$, while $F_2(\xi)$ cannot be so extended as an analytic function $F_2(\zeta)$ into $\eta > 0$ but can be so extended in $\eta < 0$; moreover for real ζ, $F(\zeta)$ is the real part of $F_1(\zeta)$ (as well as of $F_2(\zeta)$).

By applying the integral (1.7) not to F but to F_1, and taking the real part, one obtains an extension of the Weyl integral of general applicability, provided (2.10) or (2.11) applies to F_1. Actually, even this is not the case for the function F corresponding to a current pulse, and a further modification of the path of integration to handle these cases will be indicated in Section 4.

3. *Extension of Functions of a Real Variable to a Complex Variable*

The definition of the exponential integrand for complex values of the variables γ or θ in (1.1) offered no special difficulty. Similarly, no special difficulty is encountered in the interpretation of the refracted waves for the case where the angle of refraction, as given by Snell's law, has a sine greater than unity, and similar remarks apply when, as shown in Figure 2, θ is complex, since the definition of the exponential for all values of its argument, both real and complex, is clear enough. However, the interpretation of F' for complex values of its argument, when F is initially defined only for real values, is a real question that must be examined in detail.

Since (1.7) was established formally by means of Fourier integrals, these integrals will be used to furnish the key to this extension of functions of a real variable to complex values of its variable.

We consider, therefore, the extension by means of its Fourier integral of a real function $\varphi(\xi)$, defined for real ξ, to complex values of its argument obtained by replacing ξ by $\zeta = \xi + i\eta$.

Write the Fourier integral of $\varphi(\xi)$ in the form

$$\varphi(\xi) = \int_{-\infty}^{+\infty} e^{i\alpha} f(\alpha)\, d\alpha. \tag{3.1}$$

Break up this integral into two parts

$$\varphi(\xi) = \int_{0}^{\infty} e^{i\alpha\xi} f(\alpha)\, d\alpha + \int_{-\infty}^{0} e^{i\alpha\xi} f(\alpha)\, d\alpha. \tag{3.2}$$

It will be noted that if in the first integral ξ is replaced by ζ, there results

$$\int_0^{\infty} e^{i\alpha\zeta} f(\alpha)\, d\alpha = \int_0^{\infty} e^{i\alpha\xi} e^{-\alpha\eta} f(\alpha)\, d\alpha. \tag{3.3}$$

For $\eta > 0$ this is convergent and represents an analytic function of ζ in the upper half ζ-plane. A similar substitution in the second integral leads to a divergent integral. Only by replacing ξ by $\bar{\zeta} = \xi - i\eta$ will there result a convergent integral in $\eta > 0$. Hence we may extend $\varphi(\xi)$ to the half-plane $\eta > 0$ by means of

$$\varphi = \int_0^{\infty} e^{i\alpha\zeta} f(\alpha)\, d\alpha + \int_{-\infty}^{0} e^{i\alpha\bar{\zeta}} f(\alpha)\, d\alpha, \qquad \eta > 0, \tag{3.4}$$

$$\zeta = \xi + i\eta, \qquad \bar{\zeta} = \xi - i\eta.$$

Thus defined, φ is not analytic in ζ in $\eta > 0$, but is a sum of a function analytic in ζ and a function analytic in $\bar{\zeta}$. Furthermore φ is a solution of the Laplace equation

$$\frac{\partial^2\varphi}{\partial\xi^2} + \frac{\partial^2\varphi}{\partial\eta^2} = 0 \tag{3.5}$$

for $\eta > 0$, and is thus the solution of the Dirichlet problem for the boundary value $\varphi(\xi)$ for $\eta > 0$. The conditions for φ at infinity are treated by using circular inversion, say in the circle $\xi^2 + \eta^2 = 1$, and transforming the infinite region into the neighborhood of the origin.

An extension of $\varphi(\xi)$ into the region $\eta < 0$ by means of (3.4) is impossible since both integrals diverge in $\eta < 0$. In place of (3.4) one may use now

$$\varphi = \int_0^{\infty} e^{i\alpha\bar{\zeta}} f(\alpha)\, d\alpha + \int_{-\infty}^{0} e^{i\alpha\zeta} f(\alpha)\, d\alpha. \tag{3.6}$$

Remarks similar to those made about (3.4) apply to (3.6). However, the analytic function of ζ in $\eta > 0$ in the right-hand member of (3.4) does *not* continue in $\eta < 0$ into the analytic function of ζ in (3.6). Similarly, φ as defined by (3.6) is a solution of the Dirichlet problem for $\eta \leqq 0$, taking on the values φ on $\eta = 0$, but (3.4), (3.6) do *not* constitute a solution of (3.5) in the whole ζ-plane and $\partial\varphi/\partial\eta$ is discontinuous on $\eta = 0$.

In addition to the forms (3.4), (3.6) based on the Fourier integral (3.1), the function $\varphi(\xi, \eta)$ may also be represented by means of the integral

$$\varphi(\xi, \eta) = \frac{|\eta|}{\pi} \int_{-\infty}^{+\infty} \frac{\varphi(\xi')\, d\xi'}{\eta^2 + (\xi - \xi')^2}. \tag{3.7}$$

This is obtained by applying the Green's theorem to $\varphi(\xi, \eta)$ and *the* Green's function for either half plane.

based on the analyticity of F in the upper half-plane of its argument and the assumption (2.11), does not apply, in general, to cases when F is given as an arbitrary real function of its real argument.

It is shown in Section 3 that while in general it is impossible to extend an arbitrary function $F(\xi)$, defined for real values of ξ, to complex values $\zeta = \xi + i\eta$, as an analytic function of ζ over the whole ζ-plane, it is always possible to break up $F(\xi)$ into a sum of two functions

$$F(\xi) = \tfrac{1}{2}[F_1(\xi) + F_2(\xi)], \tag{2.12}$$

such that $F_1(\xi)$ can be extended as an analytic function $F_1(\zeta)$ in $\eta > 0$, while $F_2(\xi)$ cannot be so extended as an analytic function $F_2(\zeta)$ into $\eta > 0$ but can be so extended in $\eta < 0$; moreover for real ζ, $F(\zeta)$ is the real part of $F_1(\zeta)$ (as well as of $F_2(\zeta)$).

By applying the integral (1.7) not to F but to F_1, and taking the real part, one obtains an extension of the Weyl integral of general applicability, provided (2.10) or (2.11) applies to F_1. Actually, even this is not the case for the function F corresponding to a current pulse, and a further modification of the path of integration to handle these cases will be indicated in Section 4.

3. *Extension of Functions of a Real Variable to a Complex Variable*

The definition of the exponential integrand for complex values of the variables γ or θ in (1.1) offered no special difficulty. Similarly, no special difficulty is encountered in the interpretation of the refracted waves for the case where the angle of refraction, as given by Snell's law, has a sine greater than unity, and similar remarks apply when, as shown in Figure 2, θ is complex, since the definition of the exponential for all values of its argument, both real and complex, is clear enough. However, the interpretation of F' for complex values of its argument, when F is initially defined only for real values, is a real question that must be examined in detail.

Since (1.7) was established formally by means of Fourier integrals, these integrals will be used to furnish the key to this extension of functions of a real variable to complex values of its variable.

We consider, therefore, the extension by means of its Fourier integral of a real function $\varphi(\xi)$, defined for real ξ, to complex values of its argument obtained by replacing ξ by $\zeta = \xi + i\eta$.

Write the Fourier integral of $\varphi(\xi)$ in the form

$$\varphi(\xi) = \int_{-\infty}^{+\infty} e^{i\alpha} f(\alpha)\, d\alpha. \tag{3.1}$$

Break up this integral into two parts

$$\varphi(\xi) = \int_0^{\infty} e^{i\alpha\xi} f(\alpha)\, d\alpha + \int_{-\infty}^{0} e^{i\alpha\xi} f(\alpha)\, d\alpha. \tag{3.2}$$

It will be noted that if in the first integral ξ is replaced by ζ, there results

$$\int_0^\infty e^{i\alpha\zeta} f(\alpha)\, d\alpha = \int_0^\infty e^{i\alpha\xi} e^{-\alpha\eta} f(\alpha)\, d\alpha. \tag{3.3}$$

For $\eta > 0$ this is convergent and represents an analytic function of ζ in the upper half ζ-plane. A similar substitution in the second integral leads to a divergent integral. Only by replacing ξ by $\bar{\zeta} = \xi - i\eta$ will there result a convergent integral in $\eta > 0$. Hence we may extend $\varphi(\xi)$ to the half-plane $\eta > 0$ by means of

$$\varphi = \int_0^\infty e^{i\alpha\zeta} f(\alpha)\, d\alpha + \int_{-\infty}^0 e^{i\alpha\bar{\zeta}} f(\alpha)\, d\alpha, \qquad \eta > 0, \tag{3.4}$$

$$\zeta = \xi + i\eta, \qquad \bar{\zeta} = \xi - i\eta.$$

Thus defined, φ is not analytic in ζ in $\eta > 0$, but is a sum of a function analytic in ζ and a function analytic in $\bar{\zeta}$. Furthermore φ is a solution of the Laplace equation

$$\frac{\partial^2\varphi}{\partial\xi^2} + \frac{\partial^2\varphi}{\partial\eta^2} = 0 \tag{3.5}$$

for $\eta > 0$, and is thus the solution of the Dirichlet problem for the boundary value $\varphi(\xi)$ for $\eta > 0$. The conditions for φ at infinity are treated by using circular inversion, say in the circle $\xi^2 + \eta^2 = 1$, and transforming the infinite region into the neighborhood of the origin.

An extension of $\varphi(\xi)$ into the region $\eta < 0$ by means of (3.4) is impossible since both integrals diverge in $\eta < 0$. In place of (3.4) one may use now

$$\varphi = \int_0^\infty e^{i\alpha\bar{\zeta}} f(\alpha)\, d\alpha + \int_{-\infty}^0 e^{i\alpha\zeta} f(\alpha)\, d\alpha. \tag{3.6}$$

Remarks similar to those made about (3.4) apply to (3.6). However, the analytic function of ζ in $\eta > 0$ in the right-hand member of (3.4) does *not* continue in $\eta < 0$ into the analytic function of ζ in (3.6). Similarly, φ as defined by (3.6) is a solution of the Dirichlet problem for $\eta \leqq 0$, taking on the values φ on $\eta = 0$, but (3.4), (3.6) do *not* constitute a solution of (3.5) in the whole ζ-plane and $\partial\varphi/\partial\eta$ is discontinuous on $\eta = 0$.

In addition to the forms (3.4), (3.6) based on the Fourier integral (3.1), the function $\varphi(\xi, \eta)$ may also be represented by means of the integral

$$\varphi(\xi, \eta) = \frac{|\eta|}{\pi} \int_{-\infty}^{+\infty} \frac{\varphi(\xi')\, d\xi'}{\eta^2 + (\xi - \xi')^2}. \tag{3.7}$$

This is obtained by applying the Green's theorem to $\varphi(\xi, \eta)$ and *the* Green's function for either half plane.

In terms of the integrals (3.4) we now put

$$\frac{1}{2} F_1(\zeta) = \int_0^\infty e^{i\alpha\zeta} f(\alpha)\, d\alpha, \qquad \eta \geqq 0$$

(3.8)

$$\frac{1}{2} F_2(\bar{\zeta}) = \int_{-\infty}^0 e^{i\alpha\bar{\zeta}} f(\alpha)\, d\alpha, \qquad \eta \geqq 0.$$

For $\eta = 0$, $F_1(\xi)$, $F_2(\xi)$ are conjugates of each other, equation (2.12) holds. as well as the relations

(3.9) $$F(\xi) = \mathcal{R}e[F_1(\xi)] = \mathcal{R}e[F_2(\xi)].$$

In terms of the harmonic function φ defined by (3.7), we introduce F_1 by finding ψ, the conjugate harmonic to φ in $\eta > 0$ and defining F_1 as

(3.10) $$F_1(\zeta) = \varphi + i\psi.$$

4. *Extension of the Weyl Integral*

For an arbitrary real function F we now put

(4.1) $$F(\xi) = \mathcal{R}e[F_1(\xi)]$$

where F_1 is analytic in $\eta > 0$. We shall modify (1.7) as follows

(4.2) $$\frac{F(ct - R)}{R} = \mathcal{R}e\left\{-\frac{1}{2\pi}\int F_1'[ct - (\alpha x + \beta y + \gamma z)]\, d\omega\right\}.$$

If the condition (2.10) applies to F_1 :

(4.3) $$F_1(i\infty) = 0$$

then the proof outlined in equations (2.7)-(2.9), applied to the right-hand member of equation (4.2), converts it to $F_1(ct - R)/R$ and application of equation (4.1) leads to equation (4.2).

Of special interest is the case of a current pulse in the dipole, $I(t) = \delta(t)$, since by Duhamel's theorem the field due to any current shape can be resolved into a superposition of the field due to current pulses. If the left-hand member of (4.2) represents the Hertz (vector) potential then F is proportional to the dipole moment ($= \int_{-\infty}^{t} i(t')\, dt'$) and is proportional to the Heaviside unit function

(4.4) $$F(ct) = H(t) = \begin{cases} 1 & \text{for} \quad t > 0, \\ 0 & \text{for} \quad t < 0. \end{cases}$$

In this case the function F cannot be represented as a Fourier integral in view of convergence difficulties. The harmonic function $\varphi(\xi, \eta)$ is given by

(4.5) $$\varphi(\xi, \eta) = 1 - \frac{1}{\pi}\tan^{-1}(\eta/\xi) = 1 - (\arg \zeta)/\pi$$

and $F_1(\zeta)$ is readily shown to be

$$F_1(\zeta) = 1 + \frac{i}{\pi} \ln \zeta \tag{4.6}$$

while

$$F_1'(\zeta) = \frac{i}{\pi \zeta}. \tag{4.7}$$

Thus it will be seen that not only is equation (4.3) invalid, but $F_1(\zeta)$ even becomes infinite at infinity. Actually, the condition (4.3) may be replaced by the less severe condition

$$\mathcal{R}e[F_1(i\infty)] = 0, \tag{4.8}$$

but even this condition is not satisfied, the limit in question being 1/2.

To overcome the difficulty just indicated, we modify the path of the θ- and η-integrations from that shown in Figure 2 to the path shown in Figure 4 and impose the condition

$$\mathcal{R}e[F_1(-\infty)] = 0. \tag{4.9}$$

It will be noted that this condition *is* satisfied by (4.6) as well as by F_1 for cases where the dipole current vanishes for t less than fixed value t_0.

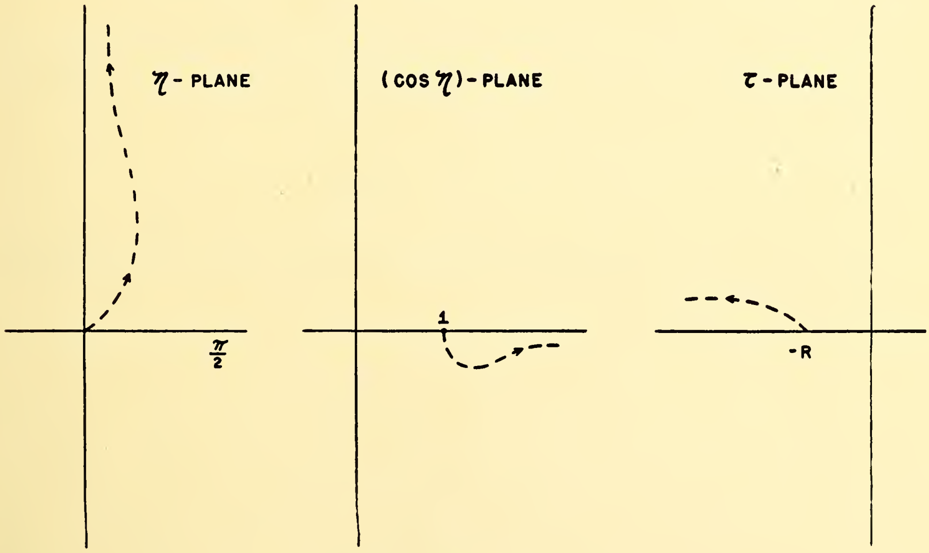


FIGURE 4

Summarizing, it has been shown that a proper extension of (1.1) applicable to arbitrary spherical waves, and especially to waves started by a current pulse, is given by (4.2) with the path of integration for θ and η as shown in Figure 4.

The integrand of (4.2) still has the advantage of being analytic in its arguments in the required domain of the latter. The interpretation in terms of plane waves is obvious, but essentially all the wave normals have complex direction cosines.

It will be noted that only if the path of θ-integration is along the solid path of Figure 2 so that γ is real or pure imaginary, will the argument

$$ct - (\alpha x + \beta y + \gamma z) = ct - \sin\theta\,(x\cos\varphi + y\sin\varphi) - z\cos\theta \tag{4.10}$$

of F_1' in (4.2), when complex, lie in the upper-half plane.

With the modified paths of Figure 4 and similar paths for θ, the quantity (4.10) no longer lies in the upper-half plane. In particular, if the θ-path is collapsed onto the pure imaginary axis, then $\cos\theta$ becomes real and >1 and $\sin\theta$ becomes positive imaginary, and (4.10) has a positive imaginary part only for certain combinations of x, y and z, but will acquire a negative imaginary part for other combinations of x, y, and z.

One is thus faced with a choice of either using the *restricted* path of integration shown as a solid line in Figure 2 and having to put up with an electrostatic term $1/2R$ arising from infinity, or else removing the restriction of confining (4.10) to the upper half-plane. The latter can be done for the impulse current when F_1 is given by (4.6), since $F_1'(\zeta)$ is analytic in the whole ζ-plane except for $\zeta = 0$. We shall adopt this second alternative, using the path of integration of Figure 4 for either θ or η, and allowing θ or η approach the pure imaginary axis. Care must be taken to avoid crossing $\zeta = 0$.

It is to be pointed out that the extension of the argument of $F_1'(\zeta)$ into $\mathscr{I}m\,(\zeta) < 0$ is not to be confused with the function $F_2'(\zeta)$ in that half-plane. Indeed, one readily shows that for the case when F_1' is given by (4.7), F_2' is the negative of F_1':

$$F_2'(\zeta) = -\frac{i}{\pi\zeta} = -F_1'(\zeta). \tag{4.11}$$

The application of the above to a dipole near a flat earth will be presented in a future paper.

Kirchhoff's Formula, Its Vector Analogue, and Other Field Equivalence Theorems

By S. A. SCHELKUNOFF
Bell Telephone Laboratories

1. *Introduction*

One of the purposes of this symposium is to discuss some of the theoretical difficulties involved in the solution of electromagnetic problems. For this reason I believe that the treatment of diffraction problems is an appropriate topic. In particular, the problem of radiation from a horn (acoustic or electromagnetic) has been treated approximately with the aid of one of several formulas suggested by Huygens' physical idea about wave propagation. These formulas are not equally powerful, do not always give equally good approximations, and do not inspire the same *a priori* confidence in the results. Of course, if no approximations were made in the formulas, it would not matter which formula was used since then the result would always be exact. However, approximations are unavoidable except when the answer is already known and when there is no need for any formula.

To understand how Huygens' physical assumption that the conditions at the front of a wave determine the subsequent wave motion suggests various formulas, let us consider a source first in an infinite homogeneous medium and then inside a perfectly rigid horn. In the first case the wavefronts are closed surfaces surrounding the source. In the second case, the wavefronts are open surfaces sliding along the walls of the horn until they reach the aperture. Eventually, the wavefronts will become closed surfaces surrounding the horn as well as the source. Kirchhoff derived an explicit formula for the field outside a closed surface (S) in terms of the wave function and its normal derivative on (S) on the assumption that the source is inside (S) and that the medium outside (S) is homogeneous. In the case of the horn Kirchhoff's surface of integration must enclose the horn as well as the source since the horn introduces a discontinuity in the medium. This closed surface may be chosen to consist of an "aperture surface" (S_a) together with the exterior surface of the horn. In the case of electromagnetic waves the present writer proved another theorem [1, 2] (the "Induction Theorem") in which the surface of integration is solely the aperture surface (S_a). An obvious analogue of this theorem for scalar waves

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

will be discussed in this paper. To understand the Induction Theorem we should note that the conditions at a wavefront before it reaches the aperture are the same whether the horn is truncated at the aperture or extended indefinitely. Hence, we might expect the existence of a formula which expresses the field of a truncated horn in terms of the field of an infinite horn. The Induction Theorem provides such a formula. The differences between Kirchhoff's formula and the Induction Formula are the following: (*1*) in the former the surface of integration must completely enclose the horn while in the latter the surface of integration is an open surface either in the interior or in the aperture but such that together with the boundary of the horn it forms a closed surface, (*2*) in Kirchhoff's formula the integrand depends on the actual field of the truncated horn while in the Induction Formula it depends on the field which exists when the horn is extended to infinity,[1] (*3*) in Kirchhoff's formula the integrand depends also on the Green's function for an infinite homogeneous medium while in the Induction Formula the Green's function is for the medium *with* the horn (that is, the Green's function must satisfy the proper boundary conditions at the surface of the horn). At first it may seem that the latter condition restricts the usefulness of the Induction Formula. This, however, is not the case. In the Induction Formula we are given a distribution of virtual sources over the aperture which in the presence of the horn produces the correct external field of the truncated horn and the field returned into the horn by the truncation. The calculation of the field of the virtual sources may be divided into two separate problems. First we can calculate the field on the assumption that the sources are in an infinite homogeneous medium; then we impress this field on the horn and try to evaluate the reflection from the horn. In the case of large apertures we shall find that the field impressed on the horn is relatively small except near the edge.

There are other formulas which express the effect of an alteration in the environment of a source (such as truncation of a horn) on the field, or the equivalence of fields produced by two different systems of sources. For example, if an electric source is inside a perfectly conducting closed surface, the field outside is zero. Hence, the currents in the surface produce an external field which is equal and opposite to that of the source. Consequently, the currents equal and opposite to those in the surface produce the same external field as the original source. This is a field analogue of Norton's circuit equivalence theorem, in which a network with interior generators is replaced by a source of fixed current in parallel with the admittance of the network. Similarly there is a field analogue of Helmholtz-Thévenin's circuit equivalence theorem in which a network with interior generators is replaced by a source of fixed voltage in series with the impedance of the network. In the field analogue certain hypothetical magnetic currents correspond to sources of fixed voltage. There are also circuit analogues of Kirchhoff's formula and of the Induction Formula.

---

[1]The actual field of the truncated horn would also give the correct result, but it would serve no useful purpose to employ it.

These various formulas are interesting from the theoretical point of view and useful from the practical point of view. Kirchhoff's formula is the oldest of all. It has been used successfully to obtain first approximations to the solutions of problems involving transmission of sound and light through large apertures in screens. However, in more recent years it has been found that attempts to apply Kirchhoff's formula to the calculation of electromagnetic radiation from horns led to various difficulties. For instance, Barrow and Greene [3] found that either the polarization or the intensity of the radiation field involved substantial errors when the formula was applied to one or another field vector. Of course, some polarization errors are also involved in optical applications; but these errors are appreciable only in directions making large angles with the principal beam, and in these directions optical fields are extremely weak. In microwave radio, on the other hand, the apertures are not particularly large compared with the wavelength, and large angles have to be considered.

There are two possible sources of error. Kirchhoff's formula requires integration of certain functions over a closed surface (S); but in applications we are forced to integrate over an "aperture surface" (S_a) which is only a small part of the total surface (S). It has been suggested [4] that having obtained the first approximation as usual we could use this approximation in Kirchhoff's formula to obtain a second approximation. But in the next section we shall presently see that the second approximation obtained in this way must equal the first and that the initial error remains intact. To resolve this difficulty we need the Induction Formula.

In the electromagnetic case, the second source of error might be in that Kirchhoff's formula involves scalar wave functions and not vector wave functions. It is true that each cartesian component of either E or H is a scalar wave function, and thus Kirchhoff's formula may be applied to it, but the various components are not independent wave functions since they must also satisfy Maxwell's equations. Thus when we apply Kirchhoff's formula to a single component and integrate over only a part of the total closed surface we may (and actually do) obtain a solution of the wave equation which does not satisfy Maxwell's equations. To overcome this difficulty, Love [5] obtained a vector analogue of Kirchhoff's formula which expresses an electromagnetic field at a given point in terms of the tangential components of E and H on a closed surface (S). Later another derivation of the same formula was given by MacDonald [6]. However, if Love's formula is applied to only a part of the closed surface (S), the result still does not satisfy Maxwell's equations. The present writer obtained [1] a simpler form of the vector analogue of Kirchhoff's formula. For a closed surface this form is mathematically equivalent to Love's form; but when our formula is applied to a portion of the closed surface, the result satisfies Maxwell's equations. Stratton and Chu resolved the same difficulty in another way [7]. When applying Love's formula to a portion of the closed surface, they add certain line integrals to the original surface integrals.

These line integrals are suggested by the physical interpretation of the integrand in Love's formula and do not follow from the mathematical proof.

However, none of these vector analogues of Kirchhoff's formula can be used in the manner suggested by Born [4] to obtain a higher order approximation. The initial error made in the integrand of either Kirchhoff's formula or its vector analogue must remain intact throughout repeated applications of the formula. To resolve this difficulty we need the Induction Theorem.

There are other difficulties inherent in Kirchhoff's formula (and its vector analogue) which are best explained by referring to its explicit analytical formulation.

2. *Kirchhoff's Formula*

There are two forms of Kirchhoff's formula, one valid only for a finite region and the other applicable to either a finite or an infinite region. First let us consider a finite region. Let u be a solution of the scalar wave equation free

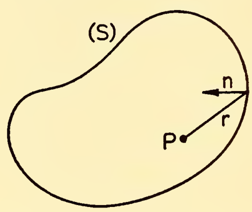


FIG. 1. A closed surface (S) and reference point P inside (S).

from singularities anywhere *inside* a closed surface (S), Figure 1. The value of u at any interior point P is determined by its value and the value of its normal derivative on (S). The best known formula is

$$(1) \qquad u(P) = \iint_{(S)} \left(u \frac{\partial}{\partial n} \frac{\exp\{-i\beta r\}}{4\pi r} - \frac{\partial u}{\partial n} \frac{\exp\{-i\beta r\}}{4\pi r} \right) dS, \qquad \beta = 2\pi/\lambda.$$

This formula is usually obtained from Green's theorem [4, 8, 9]. But this proof also gives [10]

$$(2) \qquad u(P) = \iint_{(S)} \left(u \frac{\partial}{\partial n} \frac{\cos \beta r}{4\pi r} - \frac{\partial u}{\partial n} \frac{\cos \beta r}{4\pi r} \right) dS,$$

and more generally

$$(3) \qquad u(P) = \iint \left(u \frac{\partial \psi}{\partial n} - \frac{\partial u}{\partial n} \psi \right) dS$$

where

$$(4) \qquad \psi = \frac{\cos \beta r}{4\pi r} + A \frac{\sin \beta r}{4\pi r}$$

and A is an arbitrary constant. If $A = i$, we have (1) but with advanced potentials instead of retarded potentials. What this ambiguity means is that at any *interior* point P

$$\iint_{(S)} \left(u \frac{\partial}{\partial n} \frac{\sin \beta r}{4\pi r} - \frac{\partial u}{\partial n} \frac{\sin \beta r}{4\pi r} \right) dS = 0. \tag{5}$$

The constant A does not affect the result, but, we must stress, only if we use the *exact* values of u and its normal derivative on (S). To apply these formulas to the problem of radiation from a horn we can choose (S) as shown in Figure 2,

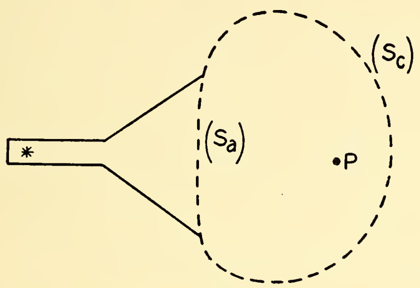


FIG. 2. A closed surface, $S_a + S_c$, surrounding reference point P outside a horn.

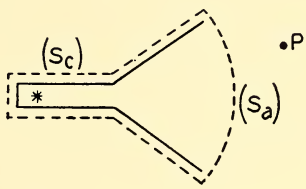


FIG. 3. A closed surface surrounding a horn with internal sources.

where (S) consists of an "aperture surface" (S_a) and a complementary surface (S_c) so that $(S_a) + (S_c)$ is a closed surface surrounding the region of interest.

We can also choose $(S) = (S_a) + (S_c)$ in such a way that it surrounds the horn completely, Figure 3. If we assume that the sources are inside the horn, we know more about their field than we did in the preceding case. Now we know not only that u is a solution of the wave equation, free from singularities

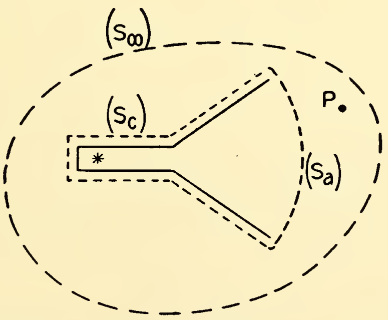


FIG. 4. Two closed surfaces, $S_a + S_c$ and S_∞, surrounding an acoustic horn.

outside (S) but also that u behaves at infinity as $r^{-1} \exp \{-i\beta r\}$. We are not permitted to apply Green's formula directly to the infinite region outside (S); but we can apply it to the finite region between (S) and another surface (S_∞)

which surrounds (S), Figure 4. As we try to remove (S_∞) to infinity we find that the contribution from this surface does not vanish *unless* the constant A in equation (2) equals $-i$. In this case we obtain equation (1). In the preceding case we could not make use of the behavior of u at infinity for the simple reason that the region of interest did not extend to infinity.

It is now clear that we should be careful how we apply Kirchhoff's formula to the problem of radiation from an acoustic horn. Sometimes the argument proceeds as follows: Choose a closed surface (S) consisting of two parts, one (S_a) in the aperture and the other (S_c) so as to enclose the point P where we wish to determine the excess pressure u. The exact values of u and $\partial u/\partial n$ are not known on (S); but if the aperture is large, then physical intuition tells us that the wave motion in the aperture is likely to be about the same, except possibly near the edge, as in a hypothetical case in which the horn is continued indefinitely. The problem of an infinite horn can be solved; hence it is assumed that we know fairly accurately u and $\partial u/\partial n$ on the aperture surface (S_a). Sometimes it is also assumed that on (S_c) both u and $\partial u/\partial n$ vanish. With these assumptions the following first approximation to $u(P)$ is obtained from (1)

$$u_1(P) = \iint_{(S_a)} \left(u_0 \frac{\partial}{\partial n} \frac{\exp\{-i\beta r\}}{4\pi r} - \frac{\partial u_0}{\partial n} \frac{\exp\{-i\beta r\}}{4\pi r} \right) dS, \tag{6}$$

where u_0 and $\partial u_0/\partial n$ are the values corresponding to the infinite horn. If we apply exactly the same arguments to equation (2), we find another first approximation

$$\bar{u}_1(P) = \iint_{(S_a)} \left(u_0 \frac{\partial}{\partial n} \frac{\cos \beta r}{4\pi r} - \frac{\partial u_0}{\partial n} \frac{\cos \beta r}{4\pi r} \right) dS. \tag{7}$$

For any number of different values of A in (4) we can obtain corresponding approximations from (3). These various approximations are far from being nearly the same; they are radically different. Experience shows that for large horns equation (6) gives very good results. Equation (7), on the other hand, gives totally wrong results. Unless we are willing to accept (6) as a semi-empirical formula, we have to revise the arguments which led to (6) and (7) as being equally permissible approximations. We can make use of the physical principle enunciated by Huygens and decide in favor of (6) on physical grounds; but this can hardly be satisfactory unless we can establish a connection between the physical principle and equation (1). The derivation of (1) from Green's formula does not establish such a connection.

Going over the argument leading to the approximation (6), we observe that it is not really necessary to assume that u and $\partial u/\partial n$ vanish over (S_c). Our practical success with this approximation means that while these values obviously do not vanish, their integrals over (S_c) must be small. Somehow we should be able to prove it. We can do this by choosing (S) as in Figure 3. Here our physical intuition tells us that the excess presure u on (S_c) is small except

near the edge. From the boundary conditions we know that the normal velocity, and therefore $\partial u/\partial n$, vanishes on (S_c). All this makes it plausible on physical grounds that (6) should be a satisfactory approximation. This is as far as we can go if we base our argument on Kirchhoff's theorem. Presently we shall see that the Induction Theorem enables us to arrive at (6) on purely mathematical grounds.

Next let us see if we can improve on the first approximation—at least in principle—by using Kirchhoff's formula (1) repeatedly. We choose a closed surface (S') very near (S) in the region to which (1) is applicable. From (6) we calculate u_1 and $\partial u_1/\partial n$ on (S'). Let us substitute these values in (1) and calculate $u_2(P)$. Is $u_2(P)$ a better approximation to $u(P)$ than $u_1(P)$? The answer is no for the simple reason that $u_2(P) = u_1(P)$. Once we have calculated $u_1(P)$ from (6), we have a solution of the wave equation which is free from singularities in the region under consideration; by Kirchhoff's formula this solution is recovered from its values and the values of its normal derivative on (S'). This argument can be presented in a different form. When we approximate equation (1) by (6) we in effect replace a system of sources (1) which is truly equivalent to the given source inside the horn by another system of sources (2) which is only approximately equivalent. Thereafter, if we apply Kirchhoff's formula to the field produced by system (2), we must recover the same field. In fact, if we could obtain increasingly better approximations by using Kirchhoff's formula repeatedly, we should find ourselves in an embarrassing position of being able to lift ourselves by our own boots. There is nothing in the formula that contains a self-correcting mechanism.

3. *Field Equivalence Theorems*

Field equivalence theorems express the identity, in certain specified regions, of the fields produced by given sources and the fields produced by appropriate sources on the boundaries of the regions. Equation (1) is within the meaning of this definition; but equation (2) is not. Since we are interested in physical applications, we shall state the theorems in physical terms; but it is not difficult to rephrase them in strictly mathematical terms. A point source will become a singularity of a specified kind. A continuous distribution of sources on a surface will become a specified discontinuity either in the wave function or in its normal derivative or both. The field "produced" by such a discontinuity is the field which has this discontinuity and which behaves at infinity as $\exp\{-i\beta r\}/r$.

In the case of acoustic waves we have two types of elementary sources: (*1*) a pair of pistons moving in opposite directions, Figure 5a, and (*2*) a single piston oscillating back and forth, Figure 5b. Across the double piston the pressure u_0 is supposed to be continuous and the particle velocity $\xi_0 = -(1/i\omega\rho_0)$ $\partial u_0/\partial n$ discontinuous by the amount $\xi^+_{0,n} - \xi^-_{0,n}$; hence, the double piston acts as a source of volume current $(\xi^+_{0,n} - \xi^-_{0,n})\, dS$, where dS is the area of one face

of the piston. Across the single piston the particle velocity is continuous but the pressure is discontinuous by the amount $(u_n^+ - u_n^-)\, dS$. In the case of electromagnetic waves the elementary sources are: (1) a surface element of

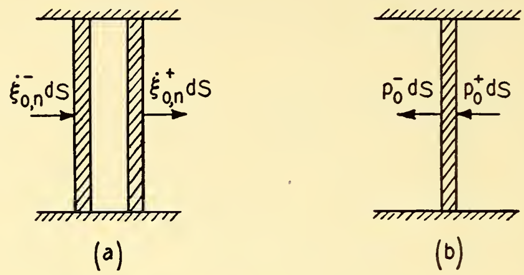


FIG. 5. Two elementary sources of acoustic waves: (a) a pair of pistons moving in opposite directions, (b) a single oscillating piston.

electric current of linear density C_e, Figure 6a, and (2) a surface element of magnetic current of linear density C_m, Figure 6b. Across the electric current

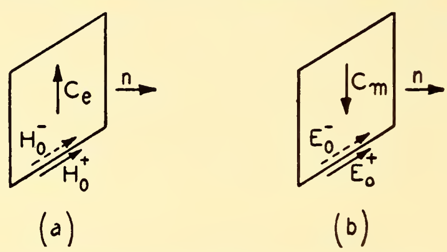


FIG. 6. Two elementary sources of electromagnetic waves: (a) a surface element of electric current, (b) a surface element of magnetic current.

element the tangential electric intensity is continuous and the tangential magnetic intensity discontinuous. The moment of the electric current element is $C_e\, dS = n(H_0^+ - H_0^-)\, dS$, where n is the normal to the element. Across the magnetic current element the tangential magnetic intensity is continuous and the tangential electric intensity is discontinuous. The moment of the magnetic current element is

$$C_m\, dS = (E_0^+ - E_0^-)n\, ds.$$

We can now state the following theorems.

4. *Acoustic Field Equivalence Theorems*

Theorem 1. *The Acoustic Induction Theorem*

Consider a rigid or a perfectly pliable horn of finite dimensions with a given system of internal sources, Figure 7a, producing a field of excess pressure u.

Let (S_a) be an "aperture surface" which together with the boundary (S_h) of the horn forms a closed surface. Let u_0 be the excess pressure produced by the same sources when the horn is continued indefinitely, Figure 7b. *The field*

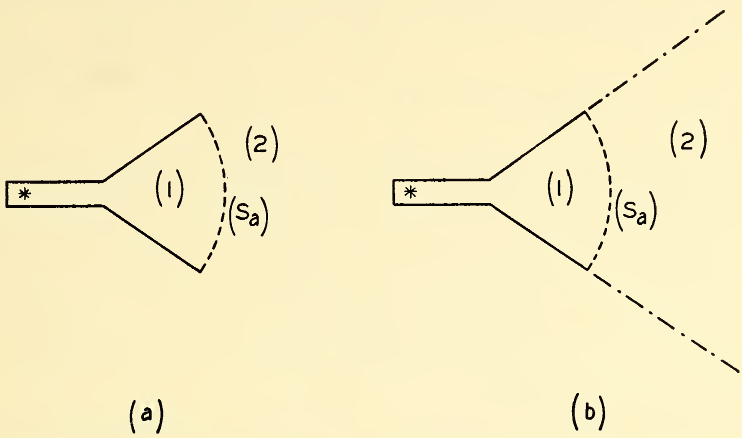


FIG. 7. A horn with internal sources: (a) horn is finite and closed through an aperture surface (S_a), (b) horn is extended indefinitely.

produced in the presence of the truncated horn by a layer of elementary sources of volume current $\xi_{0,n}\, dS = -(1/i\omega\rho_0)\, \partial u_0/\partial n$ *and a layer of elementary sources of excess pressure* $u_0\, dS$, *also on* (S_a), *equals the field* u *produced by the given sources at points external to* $(S_a) + (S_h)$. *At points internal to* $(S_a) + (S_h)$ *the field of the aperture layers equals the difference* $u - u_0$ *between the field produced by the given sources in the finite horn and the field produced by the same sources in the extended horn.*

If G is the excess pressure produced by a source of unit volume current in the presence of the given horn, then the above theorem states that

$$u(P) = \iint_{(S_a)} G i\omega\rho_0(\xi_{0,n}\, dS) + \iint_{(S_a)} \frac{\partial G}{\partial n} u_0\, dS \tag{8}$$

$$= -\iint_{(S_a)} G \frac{\partial u_0}{\partial n}\, dS + \iint_{(S_a)} \frac{\partial G}{\partial n} u_0\, dS$$

if P is an exterior point, and

$$u(P) = u_0(P) - \iint_{(S_a)} G \frac{\partial u_0}{\partial n}\, dS + \iint_{(S_a)} \frac{\partial G}{\partial n} u_0\, dS \tag{9}$$

if P is an interior point.

In particular, if the boundary of the horn is conical, both before and after its extension to infinity, and if the junction between the duct and the horn is

small compared with the wavelength, or if an acoustic lens is introduced in the junction to transform the plane equiphase surfaces in the duct into spherical in the horn, then $u_0 = A \exp\{-i\beta r_0\}/r_0$, where r_0 is the distance from the apex of the horn. Let us choose (S_a) to be a segment of the sphere of length l centered at this apex and passing through the edge of the aperture, then,

$$u(P) = -A \iint_{(S_a)} G \frac{\partial}{\partial l} \frac{\exp\{-i\beta l\}}{l}\, dS + A \iint_{(S_a)} \frac{\partial G}{\partial n} \frac{\exp\{-i\beta l\}}{l}\, dS. \tag{10}$$

Note the following differences between the Induction Theorem (8) and the Kirchhoff formula (1) as far as practical applications are concerned: (*1*) in (1) we know the Green's function $G = \exp\{-i\beta r\}/4\pi r$ but we do not know the distribution of sources on the surface of integration, while in (8) we know the distribution of sources but not the Green's function; (*2*) in (1) the surface of integration is a closed surface surrounding the given sources while in (8) it is an open surface in the aperture. In a later section we shall discuss the practical consequences of this difference.

To prove the induction theorem we start with the extended horn, Figure 7b. Let u_0 be the excess pressure produced by the given sources. Let u be the excess pressure produced by the same sources in the truncated horn, Figure 7a. We shall now consider the excess pressure given by

$$\begin{aligned} \hat{u} &= u - u_0, \qquad \text{in region (1)},\\ &= u, \qquad \text{in region (2)}. \end{aligned} \tag{11}$$

Since u and u_0 satisfy the wave equations, $\hat{u}$ also satisfies the wave equation everywhere except on the boundary (S_a) between the regions (1) and (2). At infinity $\hat{u}$ varies as $\exp\{-i\beta r\}/r$ because u varies in this manner. In region (1) $\hat{u}$ has no singularities because the sources of u and u_0 are supposed to be the same. On the surface of the horn $\hat{u}$ satisfies the required boundary conditions because u and u_0 satisfy them; across the boundary (S_a) $\hat{u}$ and $\partial\hat{u}/\partial n$ are discontinuous

$$\hat{u}^+ - \hat{u}^- = u_0, \qquad (\partial\hat{u}^+/\partial n) - (\partial\hat{u}^-/\partial n) = (\partial u_0/\partial n). \tag{12}$$

These discontinuities define the sources of $\hat{u}$ as stated in the theorem. Since a given system of sources produces a unique field, we are certain to recover $\hat{u}$ defined by (1) from the known discontinuities (12).

We can also prove this theorem as follows. Starting with the field u_0 in the extended horn, we assume that the aperture surface (S_a) is a "perfect absorber" defined in such a way that it does not affect the field in region (1) and reduces it to zero in region (2). Across (S_a) the new field is discontinuous. The increments in the excess pressure and its derivative are respectively $-u_0(S)$ and $-\partial u_0(S)/\partial n$. These discontinuities define the sources on (S_a) which together with the given sources produce u_0 in region (1) and the zero field in region (2).

The extended part of the horn may now be removed since it is in a field-free region. If we superimpose on the present field the field produced by the discontinuities $u_0(S)$ and $\partial u_0(S)/\partial n$ on (S_a), we obtain the field which has no discontinuities on (S_a). Thus we obtain the field for the truncated horn in the form given by equations (8) and (9).

While Figure 7b illustrates the "analytic" continuation of the horn and a simple type of aperture surface, the proof of the theorem holds for any continuation of the horn and any aperture surface as long as this surface joins the boundary of the horn. We require only that jointly with the boundary of the horn the aperture surface should enclose the source, Figure 8. In addition to an arbitrary continuation of the boundary of the horn we may introduce obstacles in region (2). Nevertheless, if we determine u_0 and $\partial u_0/\partial n$ and from these values obtain the equivalent layer of sources over (S_a), then this layer will produce the correct field in region (2) and the difference between the correct field and u_0 in region (1) *even after we alter region* (2) *in any way whatsoever.* We may remove obstacles altogether or in part or merely alter their shape. We may remove the obstacles and insert other obstacles. We may alter the medium itself. But we *may not* make any alterations in region (1).

Theorem 2. *Kirchhoff's Equivalence Theorem*

Consider a closed surface (S), such as $(S_a) + (S_c)$ in Figure 2 or in Figure 4, which separates the medium into two regions, one containing the sources and the horn and the other source-free. Let u be the excess pressure created by the given sources. *In the source-free region this field equals that produced by a layer of elementary sources of volume current* $\xi_n\, dS = -(1/i\omega\rho_0)\ (\partial u/\partial n)$ *on* (S) *and a layer of elementary sources of pressure* $u dS$, *also on* (S). *In the remaining region the field produced by these sources vanishes.* On account of the second part of this theorem the horn can be removed without disturbing the field of the virtua sources on (S). Therefore in computing this field we may use the Green's function for the infinite homogeneous medium; that is, we obtain equation (1) for points in the source-free region.

Proofs are analogous to those for Theorem 1. We consider the wave function $\hat{u}$ which vanishes identically in the region originally containing the given sources and the horn, and which equals u in the source-free region. Then we show this field satisfies all the requirements of the field produced by the virtual sources on (S) as defined in the theorem. Alternatively, we can postulate a perfectly absorbing layer coinciding with (S), determine the distribution of sinks on it, and then eliminate the sinks by superimposing an equal distribution of sources but with opposite signs. Thus we recover the original field.

We note that in the above proofs it makes no difference whether (S) is of the type shown in Figure 2 or in Figure 4, while the proof based on Green's formula requires separate consideration of the two cases. The present proofs are more elastic because they are independent of particular analytic expressions and can be carried through under more general conditions. On the other hand,

by the above methods we cannot derive equations (2) or (3) as easily as from Green's formula.

It may seem remarkable that the field produced by the virtual sources defined in the above theorem vanishes in the region which originally contained the horn and the given sources. A prescribed distribution of sources on a surface must produce a field which has given discontinuities in u and $\partial u/\partial n$ across the surface; but normally it is impossible to tell beforehand what values will be assumed by u and $\partial u/\partial n$ on the two sides of the surface. But under the conditions of the present theorem the sources on (S) are not arbitrary; they are obtained from the values of u and $\partial u/\partial n$ belonging to a self-consistent field possessing special properties.

Theorem 3

Consider a horn with given interior sources and an aperture surface (S_a), Figure 7a or Figure 8, which together with the boundary of the horn forms a closed surface. Let u_0 be the excess pressure for the case in which (S_a) is absolutely rigid. Then, *the excess pressure produced by the same sources in an open horn equals the sum of u_0 and the excess pressure produced by a layer of elementary*

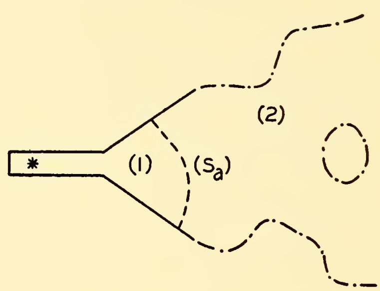


FIG. 8. An extended horn with irregular boundary and an obstacle.

sources of pressure $u_0\, dS$ distributed over (S_a). If, as in Theorem 1, G is the excess pressure produced by a source of unit volume current, then the excess pressure produced by the given source in the open horn is

$$u = u_0 + \iint\limits_{(S_a)} \frac{\partial G}{\partial n} u_0^- \, dS. \tag{13}$$

Note that each of the terms on the right is discontinuous across (S_a), but their sum is continuous. The derivative of the first term is zero on (S_a) and that of the second term is continuous.

The proof of this theorem is analogous to the proofs of the preceding theorems. We start with the field given by u_0, determine the discontinuities in u_0 and $\partial u_0/\partial n$ across (S_a), and remove them by superimposing the field with equal and opposite discontinuities.

Theorem 4

This theorem is the dual of the preceding theorem. *If u_0 is the excess pressure for the case in which (S_a) is absolutely pliable, then*

$$u = u_0 - \iint_{(S_a)} G \frac{\partial u_0^-}{\partial n}\, dS. \tag{14}$$

5. *Electromagnetic Field Equivalence Theorems*

Electromagnetic field equivalence theorems are similar to the above. Thus we have

Theorem 1. *The Electromagnetic Induction Theorem*

Consider a perfectly conducting[2] horn of finite dimensions with a given system of internal sources, Figure 7a, producing an electromagnetic field E, H. Let (S_a) be an "aperture surface" which together with the boundary (S_h) of the horn forms a closed surface. Let E_0 , H_0 be the field produced by the same sources on the assumption that the horn is extended indefinitely, Figure 7b. *The field produced in the presence of the truncated horn by an electric current sheet of density $n \times H_0$ and a magnetic current sheet of density $E_0 \times n$, both on (S_a), equals E, H at points external to $(S_a) + (S_h)$ and $E - E_0$, $H - H_0$ at the internal points.*

To prove, let

$$\begin{aligned} \hat{E} &= E - E_0\,, & \hat{H} &= H - H_0\,, & &\text{in region (1)}, \\ &= E, & &= H, & &\text{in region (2)}. \end{aligned} \tag{15}$$

Since E, H and E_0 , H_0 are solutions of Maxwell's equations, $\hat{E}$, $\hat{H}$ is also a solution. At infinity $\hat{E}$, $\hat{H}$ vary as $\exp\,\{-i\beta r\}/r$ because E, H vary in this manner. In region (1) $\hat{E}$, $\hat{H}$ has no singularities because the sources of E, H and E_0 , H_0 are supposed to be the same. On the surface of the horn $\hat{E}$, $\hat{H}$ satisfy the proper boundary conditions because E, H and E_0 , H_0 satisfy these conditions. Across the boundary (S_a) between the regions (1) and (2), the tangential components of $\hat{E}$, $\hat{H}$ are discontinuous by the amount defined in (15); thus

$$\hat{E}^+_{\tan} - \hat{E}^-_{\tan} = E_{0,\tan}\,, \qquad \hat{H}^+_{\tan} - \hat{H}^-_{\tan} = H_{0,\tan}\,. \tag{16}$$

According to Maxwell's equations these discontinuities imply magnetic and electric currents on (S_a) whose linear densities are respectively

$$C_m = E_0 \times n, \qquad C_e = n \times H_0\,. \tag{17}$$

[2]In theory the boundary could be either a perfect electric conductor defined by $E_{\tan} = 0$ or a perfect magnetic conductor defined by $H_{\tan} = 0$.

Since the field produced by a given system of sources in the presence of given boundaries is unique, we are able to recover $\hat{E}$, $\hat{H}$ from (17) as stated in the theorem.

The proof is seen to be analogous to the corresponding proof of the scalar induction theorem. Similarly the second proof of the scalar theorem can be carried over to the vector case.

Theorem 2. The Field Equivalence Theorem for Free Space

Consider a closed surface (S), such as $(S_a) + (S_c)$ in Figure 2 or Figure 3, which separates the medium into two regions, one containing the sources and the horn and the other source-free. Let E, H be the field created by the given sources. *In the source-free region this field equals that produced by an electric current sheet of density $n \times H$ and a magnetic current sheet of density $E \times n$, both on (S). In the remaining region the field produced by these sources vanishes.*

The proof is so similar to the proof already given for the acoustic case and for the electromagnetic induction theorem that we need not repeat it.

The analytic expression for the fields produced by given electric and magnetic currents may be found elsewhere [11]. We cite the results for convenience,

$$\begin{aligned} E &= -i\omega\mu A + (1/i\omega\epsilon) \operatorname{grad} \operatorname{div} A - \operatorname{curl} F, \\ H &= \operatorname{curl} A + (1/i\omega\mu) \operatorname{grad} \operatorname{div} F - i\omega\epsilon F, \end{aligned} \tag{18}$$

where

$$A = \iint_{(S)} \frac{n \times H_0}{4\pi r} e^{i\beta r}\, dS, \qquad F = \iint \frac{E_0 \times n}{4\pi r} e^{-\beta r}\, dS. \tag{19}$$

This is the form in which we expressed the vector analogue of Kirchhoff's formula (1) in our earlier papers [1, 2]. An alternative form given by Love [5], MacDonald [6], Stratton and Chu [7] is

$$\begin{aligned} E &= -\iint_{(S)} [i\omega\mu(n \times H)\psi + (n \times E) \times \nabla\psi + (n \cdot E)\nabla\psi]\, dS, \\ H &= \iint_{(S)} [i\omega\epsilon(n \times E)\psi - (n \times H) \times \nabla\psi - (n \cdot H)\nabla\psi]\, dS, \end{aligned} \tag{20}$$

where

$$\psi = e^{-i\beta r}/4\pi r. \tag{21}$$

This formula is obtained from the vector analogue of Green's formula and in the derivation for a closed surface of the type shown in Figure 1, ψ could equally well be given by equation (4). To obtain the particular form of equation (20) in which ψ is given by (21) from Green's formula we must—as in the correspond-

ing scalar case—consider two closed surfaces which enclose the horn as in Figure 4; then we can prove that for any field produced by sources inside $(S_a) + (S_c)$ the contribution from (S_∞) vanishes, provided ψ is given by (21).

All our remarks concerning Kirchhoff's scalar formula apply equally well to its vector analogue. There is one added difficulty in connection with equation (20). If (S) is replaced by the aperture surface (S_a), then E and H do not satisfy Maxwell's equations. To remedy this Stratton and Chu add certain line integrals. This solves one difficulty but raises another. What is the reason for adding these line integrals, aside from wishing to obtain expressions which satisfy Maxwell's equations? If we add arbitrarily these line integrals in the case of vector fields, why not add them also in the scalar case? From the physical point of view the line integrals are understandable, since they represent electric and magnetic charges that may be associated with the surface currents. But in the derivation of (20) from Green's formula, E and H in the integrand represent the surface values of the field intensities. More than Green's formula is needed to identify $n \times H$ and $E \times n$ with the equivalent sources on (S). Thus, in practice, we really need the stronger form of the theorem given by equations (18) and (19).

Theorem 3

Consider a perfectly conducting horn with given interior sources and an aperture surface (S_a), Figure 7a, which together with the boundary of the horn forms a closed surface. Let E_0, H_0 be the field produced by these sources when (S_a) is a perfect electric conductor. Then, *the field produced by the same sources in an open horn is the sum of E_0, H_0 and the field produced by an electric current sheet of density $n \times H_0$ over (S_a).*

This is the field analogue of Norton's circuit equivalence theorem.

The proof is analogous to that for the corresponding acoustic case (Theorem 4). The field E_0, H_0 is zero outside[3] $(S_a) + (S_h)$. Since (S_a) is a perfect conductor, $E_{\tan}$ vanishes on it and thus is continuous across (S_a). On the other hand, $H_{\tan}$ is discontinuous and the amount of the discontinuity gives the current density $-(n \times H_0)$ on (S_a). If we superimpose on E_0, H_0 the field produced by the electric current sheet of density $(n \times H_0)$ on (S_a) we eliminate the discontinuity in $H_{\tan}$ on (S_a) and maintain the continuity of $E_{\tan}$. The resultant field satisfies all the boundary conditions for the field produced by the given sources in the open horn.

Theorem 4

This is the dual of Theorem 3.

If E_0, H_0 is the field produced by the given sources when (S_a) is a perfect magnetic conductor, then the field produced by the same sources in the open horn is the

[3]It is to insure this property that we have to assume that the boundary of the horn is either a perfect electric conductor or a perfect magnetic conductor. There should be no leakage through the boundary.

sum of E_0 , H_0 *and the field produced by a magnetic current sheet of density* $E \times n$ *on* (S_a).

This is the field analogue of the Helmholtz-Thévenin circuit theorem. There is also a circuit analogue of our Induction Theorem 1 but it is neither well known nor particularly useful. In the field case, on the other hand, the analogues of the Norton and Helmholtz-Thévenin theorems are more interesting than useful while the Equivalence Theorem and the Induction Theorem acquire real importance.

6. *Possible Methods of Obtaining Higher Order Approximations to Radiation Fields of Horns*

For most practical purposes the first approximation for the radiation fields is satisfactory [12, 13, 14]. But if we are interested in the second approximation, we may be able to obtain it from the Induction Theorem. According to this theorem we have to evaluate the field of known "aperture currents" in the presence of the horn. We can consider this field as the sum of the free space field of the same currents and the field reflected from the horn. The free space field is given exactly by (19) and (20) where $(S) = (S_a)$. If we evaluate this field at the surface of the horn, particularly the tangential component of E, we shall be able to estimate the strength of the reflected field. In the case of large apertures the primary field of the aperture currents is thrown forward and the reflected field must be relatively small except near the edge of the aper-

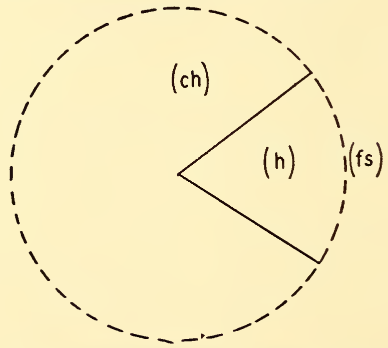


FIG. 9. A circular horn, illustrating the horn region (h), the complementary horn region (ch) and the free space region (fs).

ture. The primary field induces electric currents in the walls of the horn. In terms of the unknown density of these currents we can express their field, using (19) and (20). Equating the tangential component of E for this field to the negative of the tangential component of E for the primary field, we obtain one or two integral equations for the reflected field. Integral equations of this kind are not very tractable; but we would be trying to evaluate only a small correction term to the major part of the field given by (19) and (20).

In the case of circular horns, Figure 9, it is more practical to obtain the field of the aperture currents by successive approximations using expansions in spherical harmonics. As shown in the figure we can subdivide the entire space into three regions: the horn region (h), the free space region (fs), and the complementary horn region (ch). Assuming at first that the aperture currents are in free space, we obtain the field in all three regions. We now take the tangential field on the boundary between (fs) and (ch) and expand it into spherical harmonics appropriate to the region (ch). Similarly, we take the tangential field on the *inner* surface of the boundary between (fs) and (h) and expand it into harmonics appropriate to the region (h). Thus we obtain new fields for (h) and (ch) which satisfy proper boundary conditions on the surface of the horn. We then consider these fields as fields impressed on the free space region. In this way we obtain the second approximation to the radiation field of the horn.

Bibliography

1. Schelkunoff, S. A., *Some equivalence theorems of electromagnetics and their application to radiation problems*, Bell System Technical Journal, Volume 15, 1936, pp. 92–112.
2. Schelkunoff, S. A., *On diffraction and radiation of electromagnetic waves*, Physical Review, Volume 56, 1939, pp. 308–316.
3. Barrow, W. L., and Greene, F. M., *Rectangular hollow-pipe radiators*, Institute of Radio Engineers, Proceedings, Volume 26, 1938, pp. 1498–1519.
4. Born, M., *Optik*, Julius Springer, Berlin, 1933, p. 152.
5. Love, A. E. H., *The integration of the equations of propagation of electric waves*, Philosophical Transactions, Volume 197, Series A, 1901, pp. 1–45.
6. MacDonald, H. M., *The integration of the equations of propagation of electric waves*, London Mathematical Society, Proceedings, Volume 10, 1911, pp. 91–95.
7. Stratton, J. A., and Chu, L. J., *Diffraction theory of electromagnetic waves*, Physical Review, Volume 56, 1939, pp. 99–107.
8. Bateman, H., *Partial differential equations of mathematical physics*, University Press, Cambridge, 1932, p. 189; Dover Publications, New York, 1944.
9. Baker, B. B., and Copson, E. T., *The mathematical theory of Huygens' principle*, Oxford Press, 1939, p. 25.
10. Webster, A. G., *Partial differential equations of mathematical physics*, G. E. Stechert, New York, 1927, p. 221.
11. Schelkunoff, S. A., *Electromagnetic waves*, D. Van Nostrand, New York, 1943, pp. 128–132, eq. (1–8), (1–10), and (2–13).
12. Watson, R. B., and Horton, C. W., *The radiation patterns of dielectric rods—experiment and theory*, Journal of Applied Physics, Volume 19, 1948, pp. 661–670.
13. Watson, R. B., and Horton, C. W., *On the calculation of radiation patterns of dielectric rods*, Journal of Applied Physics, Volume 19, 1948, pp. 836–837.
14. Horton, C. W., *On the theory of the radiation patterns of electromagnetic horns of moderate flare angles*, Institute of Radio Engineers, Proceedings, Volume 37, 1949, pp. 744–749.

On the Diffraction Theory of Gaussian Optics

By H. BREMMER
Philips Research Laboratories, N. V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands

1. *Introduction*

The diffraction theory of optical systems is usually developed with reference to objects having a periodical structure or to an object consisting of a *single* point. On the other hand, when investigating the resolving power of the system it is necessary to introduce at least *two* separate point objects. For other questions, such as the connection of Abbe's intermediate image with the Fourier transform of the structure of the object, it is even necessary to consider *arbitrary* objects. The importance of considering arbitrary object structures is also evident from papers by Gabor [1] and Toraldo di Francia [2] concerning electron microscopy and phase microscopy respectively.

In this paper, too, the diffraction theory of optical imaging will be developed for objects with arbitrary structure. However, the familiar theory will here be based on rigorous solutions of the wave equation instead of the conventional approximation of Kirchhoff's formula. Apart from mathematical elegance, such a treatment is of advantage when such questions as the distribution of the wave function in the neighbourhood of the object and of its paraxial image plane are dealt with. In the first solution to be treated the similarity of the wave functions in the object plane and in the corresponding paraxial image plane (Gaussian systems with unlimited aperture) proves to be connected with Neumann's integral theorem for Bessel functions instead of the Fourier identity as in the Kirchhoff approximations. While assuming an illumination by a plane wave arriving along the optical axis of symmetry we shall discuss in succession the wave function in the object space and in the image space. It will be shown how another solution may account for the effects of optical aberrations and of limited apertures.

2. *The Transmission Function of the Object*

The illuminating primary plane wave arriving along the optical axis of symmetry (z-axis) will be given as $\exp\{i(k_0 z - \omega t)\}$. We assume the wave number k_0 with a positive (possibly infinitely small) imaginary part. The

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

complex value of k_0 is in accordance with an attenuation of the primary wave in the direction from $z = -\infty$ to $z = +\infty$ (direction of propagation); in what follows the time factor exp $\{-i\omega t\}$ will be omitted throughout.

The introduction of a single wave function does not prevent the application to vectorial problems. For such problems the wave function under discussion may represent any quantity satisfying the ordinary wave equation $(\Delta + k_0^2)u = 0$ in the space outside the imaging system. Thus our wave function can refer to any of the field components of E and H in the case of electromagnetic waves. The final differing behaviour of the various components then follows from a suitable modification of the object and of the space inside the imaging system. These modifications depend, amongst other things, on the mutual connections between these components according to Maxwell's equations. For electron optics u should be identified with Schrödinger's ψ-function.

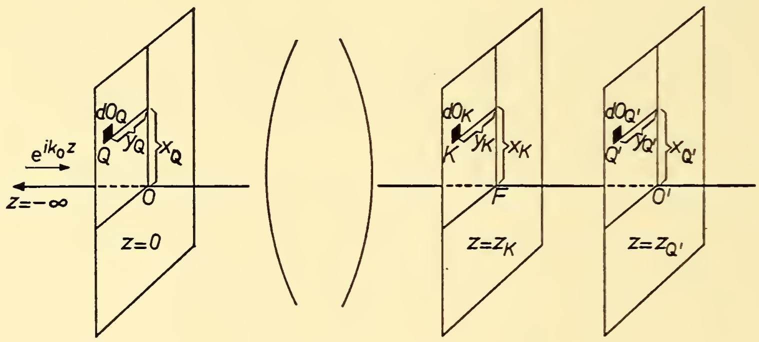


FIGURE 1

As usual we assume a two-dimensional infinitely great non reflective object at right angle to the axis of symmetry. The plane of the object will be taken as the plane $z = 0$ (see Figure 1). The primary wave exp $\{ik_0z\}$ is then propagated undisturbedly in the space $z < 0$ only and shows the constant value 1 when arriving at the object. The modification of the wave function by the object can be described by giving the distribution of the wave function directly behind the object ($z = 0+$), viz.

$$u(x, y, 0+) = 1 + \epsilon(x, y). \tag{1}$$

In general ϵ is a complex-valued function.[1] Its deviation from zero accounts for the attenuation as well as for the phase retardation caused by the object. For purely absorbing objects ϵ is real (and negative). In vectorial problems ϵ will be different for the various field components; thus it is also possible to include polarization effects due to the object. Our description is also applicable to objects of finite thickness if the plane of reference $z = 0$ is taken again directly behind the object.

[1]The function ϵ corresponds to Gabor's function $t(x, y)$; see *loc. cit.*, p. 459.

3. *The Wave Function in the Object Space*

The most important part of this space extends between the object plane and the front of the imaging system. In its region we can completely ignore the imaging system if the latter is supposed to be non reflective. The determination of the wave function u in the object space then amounts to solving the wave equation $(\Delta + k_0^2)u = 0$ under the following boundary conditions:

(a) u is equal to the given transmission function $1 + \epsilon(Q) = u(Q)$ at each point Q of the object plane,

(b) u vanishes at infinity, if $z > 0$, proportional to $\exp\{ikR\}/R$ (R = distance from the field point to the origin of the coordinate system).

This problem can be solved with Green's method which leads to the formula[2]

$$u(P) = -\frac{1}{2\pi}\frac{\partial}{\partial z_P}\iint dO_Q\, u(Q)\,\frac{\exp\{ik_0\cdot QP\}}{QP}, \tag{2}$$

where dO_Q is a surface element of the object plane and QP the distance from the object point Q to the point under consideration P. The boundary condition (1) holds according to the general identity

$$\mp\frac{1}{2\pi}\lim_{z_P\to z_{Q\pm}}\frac{\partial}{\partial z_P}\iint_{z=z_Q} dO_Q\, F(Q)\,\frac{\exp\{ik_0\cdot QP\}}{QP} = F(P), \qquad \mathcal{J}m\, k_0 \geq 0 \tag{3}$$

which has to be applied here with the upper signs when $z_Q = 0$.

Formula (3) expresses the sifting property of the following two-dimensional impulse function:

$$\mp\frac{1}{2\pi}\lim_{z\to 0\pm}\frac{\partial}{\partial z}\frac{\exp\{ik(\rho^2+z^2)^{1/2}\}}{(\rho^2+z^2)^{1/2}} = \delta(x)\,\delta(y) = \frac{\delta(\rho)}{\pi\,|\,\rho\,|}, \qquad \mathcal{J}m\, k \geq 0 \tag{4}$$

in which $\rho = (x^2 + y^2)^{1/2}$, while $(\rho^2 + z^2)^{1/2}$ is defined positive. The validity of the limit (4) (which vanishes everywhere beyond $\rho = 0$) is easily proved by an integration over ρ from 0 to ∞.

Identity (3) is of great importance in our diffraction theory and can also be considered as a geometrical interpretation of Neumann's integral theorem[3]

$$F(P) = \frac{1}{2\pi}\int_0^\infty du\, u\iint dO_Q\, F(Q)\, J_0(u\cdot PQ), \tag{5}$$

in which P is supposed to be situated in the Q-plane (a surface element of the latter is denoted by dO_Q). The equivalence of (3) and (5) is demonstrated by substituting Sommerfeld's integral for $\exp\{ik_0\cdot QP\}/QP$ in (3).

Returning to the representation (2) for the wave function in the object space we observe that this formula may be interpreted as a superposition of

[2] Compare C. J. Bouwkamp, *A contribution to the theory of acoustic radiation*, Philips Research Report 1, 1946, p. 251.

[3] See G. N. Watson, *A Treatise of Bessel Functions*, Cambridge, 1944, p. 475; compare Bouwkamp, *loc. cit.*, p. 23.

the field of dipoles situated in the object plane, the surface density of these dipoles being proportional to the transmission function $u(Q)$. Finally we emphasize the very general significance of (2). Indeed, this formula is applicable to the determination of the wave function in any half space containing no sources, if the distribution of this function has been given over the plane boundary of the half space. In other words, (2) constitutes the mathematical formulation of Huygens' principle for a non-curved surface.

4. *The Wave Function in the Image Space in Front of the Paraxial Image Plane for Gaussian Systems with Unlimited Aperture*

The wave function in the image space (the space beyond the imaging system) depends, amongst other things, on the properties of the imaging system. These properties are always given in terms of geometrical optics and therefore enable us to construct a geometric-optical approximation. In the case of a Gaussian system this approximation proves to be a rigorous solution of the wave equation and can be taken as the final solution in the case of an unlimited aperture.

In the above theory we considered a two-dimensional object. We can imagine geometric-optical rays diverging from each object point Q, these rays being transformed into convergent beams by the imaging system. We have to consider the rays leaving a single object point first. Moreover we confine ourselves provisionally to an infinitely small pencil of the rays radiating from such an object point Q. The cross-section of such a pencil changes along its trajectory, the variation being connected with that of the wave function. This follows from the law of conservation of energy. Indeed, in isotropic media the vector of the energy current density is directed along the tangents of the trajectory so that the energy radiated at Q into the pencil can never escape from it. On the other hand, the amplitude of the energy current density is proportional to the square of the modulus $|u|$ of the wave functions satisfying the scalar wave equation. The conservation of the energy current through the pencil under consideration is thus expressed by the condition

$$|u|^2 \, d\sigma = \text{constant} \tag{6}$$

along the entire trajectory.

The above considerations apply to any optical systems; now we concentrate on Gaussian systems. Here a point source at Q in the object space corresponds in the image space to a point source at the paraxial image point Q' of Q. Special normalization thus leads to the combination:

$$\frac{\exp\{ik_0 \cdot QP\}}{QP}, \qquad P \text{ in object space,} \tag{7a}$$

$$A\,\frac{\exp\{-ik_0 \cdot Q'P\}}{Q'P}, \qquad P \text{ in image space;} \qquad z_P < z_{Q'}\,. \tag{7b}$$

The different signs in the exponents are in accordance with the propagation *away* from Q in the object space and *towards* Q' in the first half of the image space. The value of $|A|$ follows at once from a comparison of the situation at the two points of intersection P_1 and P_2 of the above pencil with the two principal planes of the imaging system (see Figure 2). The Gaussian properties of the

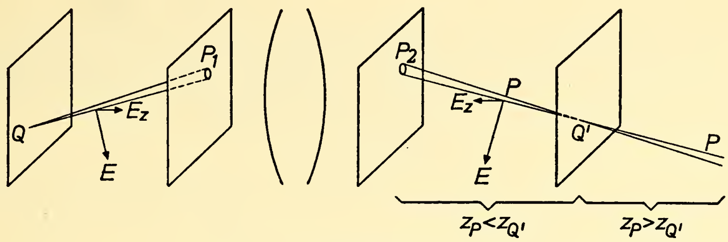


FIGURE 2

latter involve $d\sigma_{P_1} = d\sigma_{P_2}$ (in the usual approximations of the paraxial theory, which neglects the inclination of the pencil with respect to the axis of symmetry). Next we infer from (6) that $|u(P_1)| = |u(P_2)|$, or in virtue of (7)

$$|A| = Q'P_2/QP_1 .$$

This ratio represents the paraxial magnification N (again neglecting the inclination of the rays) so that $|A| = N$. It is reasonable to assume (7b) with a phase factor in accordance with the concept of an optical path length increasing continuously along the ray-trajectory QP. This phase factor proves to be exp $\{ik_0 \cdot \overline{QP}\}$, $\overline{QP}$ being defined as the optical distance:

$$\frac{1}{k_0} \int_Q^P k(s)\, ds,$$

while k is different from k_0 inside the imaging system only. The proportionality of (7b) with exp $\{ik_0 \cdot \overline{QP}\}$ involves a factor exp $\{ik_0(\overline{QP} + \overline{PQ'})\}$ = exp $\{ik_0\, \overline{QPQ'}\}$ which is to be contained in A. Evidently, the total optical path length $\overline{QPQ'}$ is independent of P, all the rays connecting Q and Q' having the same optical length (Fermat's principle) so that $\overline{QPQ'}$ may simply be marked as $\overline{QQ'}$.

Independent of the phase changes connected with the optical distance, the wave function may change its sign in some point (in the imaging system) which corresponds to a zero amplitude of this function. An additional phase factor exp $\{i\pi\} = -1$ then has to be added. This happens for instance for the z-component of the E vector in the case of a converging optical lens. In fact, the transversality and continuity of E (see Figure 2) imply opposite signs of E_z, apart from the changes due to the phase factor exp $\{ik_0 \cdot \overline{QP}\}$, at either side of the optical system.

Remembering that $|A| = N$, we thus arrive at two possible values of A, viz. $\pm N$ exp $\{ik_0 \cdot \overline{QQ'}\}$. The solution (7) may accordingly be replaced by

$$\frac{\exp\{ik_0 \cdot QP\}}{QP}, \qquad P \text{ in object space}; \qquad z_P > 0,$$

$$\text{(8)} \qquad \pm N \frac{\exp\{ik_0 \cdot \overline{QP}\}}{Q'P} = \pm N \exp\{ik_0 \cdot \overline{QQ'}\} \frac{\exp\{-ik_0 \cdot Q'P\}}{Q'P},$$

$$P \text{ in image space}; \qquad z_P < z_{Q'} .$$

We return to formula (2) for the complete wave function in the object space. Leaving out of consideration the operator $\partial/\partial z_P$, the amplitude of the contributions due to the point source at Q is given by

$$-(1/2\pi)\, dO_Q\, u(Q).$$

According to the above considerations this corresponds to a contribution of

$$\mp \frac{N}{2\pi} dO_Q\, u(Q) \frac{\exp\{ik_0 \cdot \overline{QP}\}}{Q'P}$$

for the wave function in the image space $z_P < z_{Q'}$. A final integration over the object plane while also taking into account the operator $\partial/\partial z_P$ yields the following expressions for the wave function in the first part of the image space (for an object at $z = 0$ illuminated by the plane wave $\exp\{ik_0 z\}$):

$$\text{(9a)} \qquad u(P) = \mp \frac{N}{2\pi} \frac{\partial}{\partial z_P} \iint dO_Q\, u(Q), \frac{\exp\{ik_0 \cdot \overline{QP}\}}{Q'P}$$

$$\text{(9b)} \qquad = \mp \frac{N}{2\pi} \frac{\partial}{\partial z_P} \iint dO_Q\, u(Q) \exp\{ik_0 \cdot \overline{QQ'}\} \frac{\exp\{-ik_0 \cdot Q'P\}}{Q'P},$$

$$z_P < z_{Q'} .$$

With respect to this formula we would make the following remarks:
(a) (9) represents a rigorous solution of the wave equation, being the superposition of a number of dipoles situated at the various points Q'. Therefore we may consider (9) to be the exact expression for a Gaussian system with unlimited aperture;
(b) the parameter N has been put before the integration sign. This is allowed because the paraxial theory involves a magnification that is independent of the special situation of the point Q in the object plane;
(c) the sign of (8) is taken to be the same for all object points Q, so that this very sign could be put in front of (9). This involves that the wave function under consideration should have a zero point either on any trajectory connecting the finite object with P or on none of these trajectories. This condition is satisfied for the field components usually considered.

5. *The Wave Function in and beyond the Paraxial Image Plane (Unlimited Aperture)*

The wave function in these parts of the image space has to be a continuation of the expressions (9). Formula (9b) is particularly suited for the determination

of u in the paraxial image plane $z = z_{Q'}$ itself when the integration is extended over this plane instead of over the object plane $z = 0$. This change of the domain of integration is effected by the transformation

$$x_Q = \frac{x_{Q'}}{N}, \qquad y_Q = \frac{y_{Q'}}{N}$$

which indicates that the distances between the points Q' are N times larger than those between the corresponding points Q (see also Figure 1). At the same time we obtain the new surface element

$$dO_{Q'} = dx_{Q'}\, dy_{Q'} = N^2\, dx_Q\, dy_Q = N^2\, dO_Q\,.$$

In this way (9b) is transformed into

$$(10) \qquad u(P) = \mp \frac{1}{2\pi N} \frac{\partial}{\partial z_P} \iint dO_{Q'} \times u(Q) \exp\{ik_0 \cdot \overline{QQ'}\} \times \frac{\exp\{-ik_0 \cdot Q'P\}}{Q'P}, \qquad z_P < z_{Q'}\,.$$

Now let P approach some point P_0 of the image plane $z = z_{Q'}$. The value of $u(P_0)$ then follows from formula (3) in which we should apply the *lower* signs because P is approaching the plane of integration from the left. Moreover k_0 has to be replaced by $-k_0$ in (3), which is permissible for real k_0. Evidently the function F to be substituted in (3) is given by

$$F(P_0) = \mp \frac{1}{N} u(P_0') \exp\{ik_0 \cdot \overline{P_0'P_0}\},$$

P_0' indicating the object point having P_0 as paraxial image. The final result of the limiting procedure becomes

$$(11) \qquad u(P_0) = \mp \frac{1}{N} u(P_0') \exp\{ik_0 \cdot \overline{P_0'P_0}\}, \qquad z_{P_0} = z_{Q'}\,.$$

The similarity of the distributions of the wave function in the object plane and in the conjugated image plane here results from Neumann's integral theorem for Bessel functions which is essentially identical with formula (3).

As for the space beyond the paraxial image plane, (9a) proves to be still valid *if its sign is inverted.* This can be verified as follows. For $z_P > z_{Q'}$, $\overline{QP} = \overline{QQ'} + Q'P$ holds as will be clear from the fact that, on its way, the ray connecting Q and P has to pass necessarily along Q'. The transformation of (12a) into an integral over the plane $z = z_{Q'}$, then leads to the following expression instead of (10):

$$u(P) = \pm \frac{1}{2\pi N} \frac{\partial}{\partial z_P} \iint dO_{Q'} \times u(Q) \exp\{ik_0 \cdot \overline{QQ'}\} \times \frac{\exp\{ik_0 \cdot Q'P\}}{Q'P}, \qquad z_P > z_{Q'}\,.$$

Point P may now approach the plane $z = z_{Q'}$ from the back. This implies an application of (3) with the *upper* signs which leads once more to (11). Summarizing we have the following formulae in the image space with integrations extending over the object plane:

$$
u(P) = \begin{cases} \mp \dfrac{N}{2\pi} \dfrac{\partial}{\partial z_P} \displaystyle\iint dO_Q\, u(Q) \dfrac{\exp\{ik_0 \cdot \overline{QP}\}}{Q'P}, & z_P < z_{Q'} \\[2ex] \mp \dfrac{1}{N} u(P') \exp\{ik_0 \cdot \overline{P'P}\}, & z_P = z_{Q'} \\[2ex] \pm \dfrac{N}{2\pi} \dfrac{\partial}{\partial z_P} \displaystyle\iint dO_Q\, u(Q) \dfrac{\exp\{ik_0 \cdot \overline{QP}\}}{Q'P}, & z_P > z_{Q'}\,. \end{cases} \tag{12}
$$

It is remarkable that the continuous wave function in the image space has to be represented by two different formulae at either side of the image plane. For a single object point (corresponding to $u(Q) = \delta(x_Q - x_0)\delta(y_Q - y_0)$) the different sign can be interpreted as a phase-shift π accompanying the passing along the image point, i.e. along the focal point of the beam in the image space. The attention to phase shifts of this type has been drawn by Debye.

6. *Derivation of a Rigorous Solution of the Wave Equation for Gaussian Systems with Limited Aperture*

The remark made at the end of Section 3 with respect to Huygens' principle applies to the image space as well as to the object space because the wave equation is valid in either of them. Thus we can also compute a wave function in the image space beyond some plane $z = z_K$ by substituting in (2) the wave function distribution $u(K)$ instead of $u(Q)$, K being an arbitrary point of $z = z_K$. This procedure enables us to account for the effect of an aperture in a plane perpendicular to the symmetry-axis. While identifying this plane with $z = z_K$, we take $u(Q)$ equal to the undisturbed wave function inside the aperture, and equal to zero beyond it. Therefore we can substitute (9b) for $u(K)$ inside the boundary of the aperture, because the latter is always situated between the optical system and the paraxial image plane.

In this way we get the four-dimensional integral

$$
\begin{aligned} u(P) = {} & \pm \frac{N}{4\pi^2} \iint dO_K \frac{\partial}{\partial z_P} \frac{\exp\{ik_0 \cdot KP\}}{KP} \\ & \cdot \iint dO_Q\, u(Q) \exp\{ik_0 \overline{QQ'}\} \frac{\partial}{\partial z_K} \frac{\exp\{-ik_0 \cdot Q'K\}}{Q'K}, \qquad z_K < z_P\,; z_K < z_{Q'}\,, \end{aligned} \tag{13}
$$

dO_K being a surface element of the K-plane, the integration of which has to be stopped at the boundary of the aperture.

Contrarily to (12), the solution (13) is represented by a *single* formula for the entire image space beyond the plane $z = z_K$. Either solution satisfies the rigorous wave equation and accounts for the geometric-optical properties of the Gaussian system.

7. *Modification of the Wave Function in Some Intermediate Plane*

The derivation of (13) is particularly important because it can be extended so as to include also the effects of artificial modifications imposed on the diffrac-

tion pattern in some plane perpendicular to the axis. This intermediate plane may coincide with the above aperture plane $z = z_K$ which we still assume to be situated between the imaging system and the image plane. The above modifications can be described by a function $\varphi(K)$ the modulus and phase of which correspond to the attenuation and phase retardation effected in the K-plane. The function $\varphi(K)$ may be compared with the transmission function of the object (see Section 2). Evidently $\varphi(K)$ has to be added as a factor to the K-integrand of (13) in order to account for the modifications under consideration. Moreover, we are free to invert the order of integrations in this integral; thus we get the two following identical extensions of (13):

$$u(P) = \pm \frac{N}{4\pi^2} \iint dO_K \, \varphi(K) \frac{\partial}{\partial z_P} \frac{\exp\{ik_0 \cdot KP\}}{KP} \cdot \iint dO_Q \, u(Q) \exp\{ik_0 \cdot \overline{QQ'}\} \frac{\partial}{\partial z_K} \frac{\exp\{-ik_0 \cdot Q'K\}}{Q'K}, \tag{14a}$$

$$u(P) = \pm \frac{N}{4\pi^2} \iint dO_Q \, u(Q) \exp\{ik_0 \cdot \overline{QQ'}\} \iint dO_K \, \varphi(K) \frac{\partial}{\partial z_P} \frac{\exp\{ik_0 \cdot KP\}}{KP} \frac{\partial}{\partial z_K} \frac{\exp\{-ik_0 \cdot Q'K\}}{Q'K}, \quad z_K < z_P \, ; z_K < z_{Q'} \, . \tag{14b}$$

The physical interpretation of the inversion of the order of integration in similar expressions has been emphasized by Zernike [3]. In our case the inner integral of (14a) represents the diffraction pattern established in the K-plane by the *complete* object. Therefore we can interpret (14a) as the imaging of this diffraction pattern. In (14b), however, the imaging of a *single* object point constitutes the starting point, the final integration referring to the superposition of the effects of the individual object points.

In practice, for instance in phase microscopy, the focal plane in the image space is often used as intermediate plane. The inner integral of (14a) or (13) then represents the well-known intermediate image of Abbe if the artificial modifications are absent ($\varphi(K) \equiv 1$ inside the aperture). Further, the effect of the aperture itself is included in (14) by considering the integration as extending over the complete K-plane and taking $\varphi(K)$ equal to 1 inside the aperture and equal to zero beyond it. Thus a circular aperture with radius a and centre on the optical axis is described by

$$\varphi(K) = U(a^2 - x_K^2 - y_K^2),$$

U being the unit function ($U(x) = 1$ for $x > 0$ and $U(x) = 0$ for $x < 0$). Even the theory of optical aberrations can be included by the introduction of a similar function φ, which, however, in that case in general depends on both K and Q (compare Section 9).

The practical significance of (14b) may be illustrated by the example of a small absorbing phase plate as used in phase microscopy. If we assume this

plate to be infinitely small and situated at the point K_0 on the axis of symmetry, the modification is given by

$$\varphi(K) = 1 + iA\,\delta(x_K)\,\delta(y_K).$$

Substitution of this in (14b) yields the ordinary undisturbed wave (corresponding to $\varphi = 1$) and a secondary wave given by

$$(15)\quad \mp \frac{iAN}{4\pi^2} \frac{\partial}{\partial z_P} \frac{\exp\{ik_0 \cdot K_0P\}}{K_0P} \iint dO_Q \cdot u(Q) \exp\{ik_0 \cdot \overline{QQ'}\} \frac{\partial}{\partial z_{Q'}} \frac{\exp\{-ik_0 \cdot Q'K_0\}}{Q'K_0}.$$

This wave can be interpreted as arising from a dipole at K_0. If K_0 coincides with the focal point F_2 in the image space, we can, moreover, apply the relation

$$(16)\qquad \overline{QQ'} = \overline{OF_2} + F_2Q',$$

which follows from a consideration of the trajectory QF_2Q' and in which $\overline{OF_2}$ is the optical distance along the axis between object plane and focal plane. Evaluation of (15) with the aid of (16) leads to the following simplified expression for the secondary wave generated by the phase plate in the focal point:

$$\mp \frac{AN}{4\pi^2} k_0 \cdot F_2O' \cdot \exp\{ik_0 \cdot \overline{OF_2}\} \frac{\partial}{\partial z_P} \frac{\exp\{ik_0 \cdot F_2P\}}{F_2P} \iint dO_Q\, u(Q) \frac{1 + \dfrac{1}{ik_0 \cdot F_2Q'}}{F_2Q'^2}.$$

When approximating F_2Q' by F_2O' we find the well-known proportionality of this wave with the average transmission of the object.

8. *Transition to the Usual Approximations*

Once more we assume the intermediate or aperture plane to be identical with the focal plane ($z = z_F$). The usual approximation for the wave function in the image space is then obtained from (14b) by applying the following substitutions based on the assumptions that the distances KP and KQ' are large compared to the wavelength and that the inclinations of KP and KQ' with respect to the axis of symmetry are small:

(a) ik_0 for $\partial/\partial z_P$ and $\partial/\partial z_K$;

(b) $z_P - z_F$ and $z_{Q'} - z_F$ for KP and KQ' respectively in the denominators;

$$(c)\qquad KP \sim z_P - z_F + \frac{(x_P - x_K)^2 + (y_P - y_K)^2}{2(z_P - z_F)},$$

and the corresponding expression for KQ';

$$(d)\qquad \overline{QQ'} \sim \overline{OO'} + \frac{x_{Q'}^2 + y_{Q'}^2}{2(z_{Q'} - z_F)},$$

$\overline{OO'}$ being the optical distance along the axis between object plane and image plane, and f the focal length.

Moreover we apply the following relations of Gaussian optics:

$$x_{Q'} = Nx_Q \; ; \qquad y_{Q'} = Ny_Q \; ; \qquad z_{Q'} - z_F = -Nf.$$

The final result,

$$u(P) = \mp \frac{Nk_0^2 \exp\left\{ik_0\left[\overline{OO'} + z_P - z_{Q'} + \dfrac{x_P^2 + y_P^2}{2(z_P - z_F)}\right]\right\}}{4\pi^2(z_P - z_F)(z_{Q'} - z_F)} \iint dO_Q \, u(Q) \iint dO_K$$

$$\cdot \varphi(K_x) \exp\left\{-ik_0\left[\left(\frac{x_Q}{f} + \frac{x_P}{z_P - z_F}\right)x_K + \left(\frac{y_Q}{f} + \frac{y_P}{z_P - z_F}\right)y_K\right.\right.$$

$$\left.\left. - \frac{1}{2}\left(\frac{1}{Nf} + \frac{1}{z_P - z_F}\right)(x_K^2 + y_K^2)\right]\right\}$$

then represents approximately the wave function in the image space (for a Gaussian system) corresponding to an object at $z = 0$ and an illuminating beam given by $\exp\{ik_0 z\}$. The function $\varphi(K)$ describes modifications imposed in the focal plane of the image.

This general formula can be simplified considerably for P in the paraxial image plane ($z_P = z_{Q'}$), the simplification being based on the cancelling of the quadratic terms in the exponent in the integrand. In this case the final expression becomes

$$u(P) = \mp \frac{k_0^2 \exp\{ik_0\overline{P'P}\}}{4\pi^2 f^2 N} \iint dO_Q \, u(Q) \iint dO_K \, \varphi(K)$$

(17)

$$\cdot \exp\left\{\frac{ik_0}{Nf}\left[(x_P - Nx_Q)x_K + (y_P - Ny_Q)y_K\right]\right\}.$$

For $\varphi(K) = 1$ (unlimited aperture) formula (17) reduces to a Fourier integral resulting in expression (11) which was derived above from the solution (10). Because we started in this section from solution (14) or (13), the equivalence of either solution is confirmed here in the case of an unlimited aperture as far as the usual approximations are concerned.

9. *Extension to Non-Gaussian Systems*

For these systems too we can derive a geometric-optical expression for the wave function in the image space by applying the arguments of section 4. Contrarily to Gaussian systems, however, this expression constitutes no solution of the wave equation. Nevertheless it can be considered the saddle-point approximation of such a solution as will be discussed now.

We again start from the point-source solution $\exp\{ik_0 \cdot QP\}/QP$ in the object space. The modulus $|u|$ of the corresponding geometric-optical approximation in the image space has to be derived from formula (6) which concerned

the conservation of the energy current inside a narrow pencil of trajectories. We call $d\sigma_P$ a surface element at P of the wave front passing through this point. This surface element constitutes also the cross-section of an infinitely small pencil of rays leaving Q. The spatial angle $d\Omega_i$ of these rays is constant throughout the trajectory across the image space as follows from the rectilinear course of the rays. Considering the properties of the Gaussian curvature $K(P)$ of the wavefront at P we derive

$$d\sigma_P = \frac{d\Omega_i}{K(P)} \cdot$$

Calling $d\Omega_\sigma$ the constant value of the spatial angle of the pencil under consideration in the object space we deduce from (6) by comparing the values of $| u |^2 \, d\sigma$ in the object space and at P:

$$\frac{1}{QP^2} \times QP^2 \, d\Omega_\sigma = | u(P) |^2 \frac{d\Omega_i}{K(P)} \cdot$$

Hence:

$$| u(P) | = \left[K_Q(P) \left(\frac{d\Omega_\sigma}{d\Omega_i} \right)_{Q,P} \right]^{1/2};$$

the index Q indicates reference to a special object point.

We choose the phase of u in accordance with the optical distance $\overline{QP}$ from Q to P. Thus we arrive at the following geometric-optical approximation which replaces formulae (7) and (8) for Gaussian systems:

$$\text{(18a)} \qquad \frac{\exp \{ik_0 \cdot QP\}}{QP}, \qquad P \text{ in object space}$$

$$\text{(18b)} \qquad \pm \left[K_Q(P) \left(\frac{d\Omega_0}{d\Omega_i} \right)_{Q,P} \right]^{1/2} \exp \{ik_0 \cdot \overline{QP}\}, \qquad P \text{ in image space.}$$

The expression (18b) proves to be the saddle-point approximation of the integral

$$\text{(19)} \qquad J = \pm \frac{ik_0}{2\pi} \exp \{ik_0 \cdot H(Q)\} \iint_{W(Q)} [d\Omega_\sigma \, d\Omega_i]^{1/2} \exp \{ik_0 \cdot PP_W\}$$

the integration of which extends over an arbitrary wave front in the image space for the trajectories leaving Q; PP_W represents the distance from P to the tangent plane of this wave front in any of its points W (see Figure 3), $H(Q)$ the optical distance from Q to an arbitrary point W. The verification of (18b) as the saddle-point approximation of (19) is facilitated by the introduction of a rectangular co-ordinate system $\xi\eta\zeta$ with its origin at the point of intersection P_0 of the ray QP with the wave front. We take the ζ-axis along P_0P, and the ξ-axis and η-axis along the tangents of the lines of curvature passing through P_0. As to the curvature properties we can replace the wave front near P_0 by an enveloping quadric which has the following equation:

$$\zeta = -\frac{1}{2} \left(\frac{\xi^2}{\rho_1} + \frac{\eta^2}{\rho_2} \right),$$

ρ_1 and ρ_2 being the radii of curvature at P_0. In the $\xi\eta\zeta$ system the distance PP_W reads:

$$PP_W = PP_0 - \frac{PP_0 + \rho_1}{2\rho_1^2}\,\xi_W^2 - \frac{PP_0 + \rho_2}{2\rho_2^2}\,\eta_W^2\,,$$

from which we deduce the saddle-point values $\xi_W = \eta_W = 0$ for the exponential factor in (19). In other words, the saddle point coincides with P_0. The integral (19) now reduces to (18b) if the expression

$$\frac{[d\Omega_\sigma\, d\Omega_i]^{1/2}}{d\xi\, d\eta} = K(W)\left[\frac{d\Omega_\sigma}{d\Omega_i}\right]^{1/2} \frac{d\sigma_P}{d\xi\, d\eta}$$

is replaced by its saddle-point value

$$K(P_0)\left[\left(\frac{d\Omega_\sigma}{d\Omega_i}\right)_{P_0}\right]^{1/2} = \frac{1}{\rho_1\rho_2}\left[\left(\frac{d\Omega_\sigma}{d\Omega_i}\right)_{P_0}\right]^{1/2}$$

at any point W. The verification still depends on the geometrical relation

$$K_Q(P) = \frac{1}{(P_0P + \rho_1)(P_0P + \rho_2)}\,,$$

which holds according to the generation of the wavefront through P by a shift of the points W over a distance P_0P along the normals (see Figure 3).

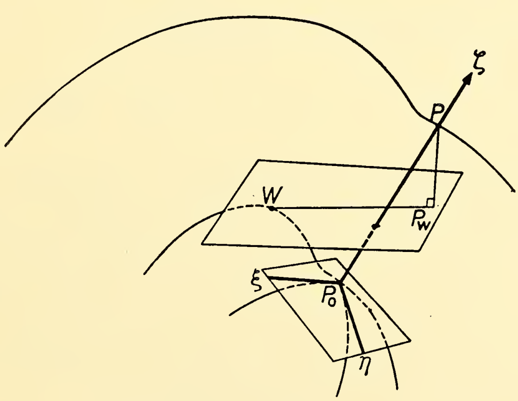


FIGURE 3

The integral (19) is composed of plane waves with wave number k_0 and thus determines a rigorous solution of the wave equation. Therefore we replace (18) by the following corresponding solutions in object space and image space:

$$(20)\quad \frac{\exp\{ik_0\cdot QP\}}{QP}\,, \qquad \pm\frac{ik_0}{2\pi}\exp\{ik_0\cdot H(Q)\}\iint\limits_{W(Q)} [d\Omega_\sigma\, d\Omega_i]^{1/2}\exp\{ik_0\cdot PP_W\},$$

the saddlepoint approximations of which satisfy the necessary geometric-optical conditions.

Returning to our problem of an object at $z = 0$ illuminated by the plane

wave $\exp\{ik_0 z\}$, we still have to apply the operator $-(1/2\pi)(\partial/\partial z_P) \iint dO_Q\, u(Q) \cdots$, in accordance with (2). Thus we get the following rigorous solution in the image space for the diffraction field due to an object illuminated by a plane wave arriving along the axis:

$$u(P) = \mp \frac{ik_0}{4\pi^2} \frac{\partial}{\partial z_P} \iint dO_Q\, u(Q) \exp\{ik_0 \cdot H(Q)\} \tag{21a}$$

$$\cdot \iint_{W(Q)} [d\Omega_\sigma\, d\Omega_i]^{1/2} \exp\{ik_0 \cdot PP_W\}.$$

Besides, we have the less accurate formula based on (18b) instead of (20):

$$u(P) \sim \mp \frac{1}{2\pi} \frac{\partial}{\partial z_P} \iint dO_Q\, u(Q) \left[K_Q(P) \left(\frac{d\Omega_\sigma}{d\Omega_i} \right)_{Q,P} \right]^{1/2} \exp\{ik_0 \cdot \overline{QP}\}. \tag{21b}$$

As in section 7 we consider an intermediate plane $z = z_K$. The distribution of the wave function existing in that plane according to (21) may be written formally as

$$u(K) = \mp \frac{N}{2\pi} \frac{\partial}{\partial z_K} \iint dO_Q\, u(Q) \varphi(Q, K) \frac{\exp\{ik_0(\overline{QQ'} - KQ')\}}{KQ'} \tag{22}$$

if $\varphi(Q, K)$ has been defined; e.g. when applying (21b)

$$\varphi(Q, K) = KQ'[K_Q(K)]^{1/2} \times \frac{1}{N} \left[\left(\frac{d\Omega_\sigma}{d\Omega_i} \right)_{Q,K} \right]^{1/2} \tag{23}$$

$$\times \exp\{ik_0(\overline{QK} - \overline{QQ'} + KQ')\}.$$

According to formula (2) we arrive again at expressions (14) for the field beyond the intermediate plane if $\varphi(K)$ has been replaced by $\varphi(Q, K)$. This function, however, must always be put in the inner integral owing to its dependence on both Q and K. Of course $\varphi(Q, K)$ should reduce to unity for a Gaussian system with unlimited aperture. In the case of the geometric-optical approximations this is evident from the three factors indicated in (23). These factors account for the non-Gaussian deviations connected in succession with:

(a) the curvature of the wavefronts in the image space,
(b) the ratio of the spatial angles of one and the same pencil in object space and image space.
(c) the optical distance from Q to the point K in the intermediate plane.

Bibliography

1. Gabor, D., *Microscopy by reconstructed wave-fronts*, Proceedings of the Royal Society of London, Series A, Volume 197, 1949, p. 454.
2. Toraldo di Francia, G., *Il fenemeno di Gibbs nella microscopia in contrasto di fase*, Il Nuovo Cionento, Volume 6, 1949, p. 30.
3. Zernike, F., *Phase contrast, a new method for the microscopic observation of transparent objects*, Physica, Volume 9, 1942, p. 686.

Diffraction and Reflection of Pulses by Wedges and Corners*

By JOSEPH B. KELLER and ALBERT BLANK
Mathematics Research Group,
Washington Square College,
New York University

1. *Introduction*

The diffraction and reflection by a perfectly conducting wedge of a periodic plane wave (with wavefront parallel to the edge) has been treated by MacDonald [1], who obtained a series of Bessel functions for the solution. We will consider the corresponding problem for an incident plane *pulse*. This problem could be solved by employing a Fourier Integral of MacDonald's solution. However, the solution can be obtained directly as an explicit closed expression in terms of elementary functions. This is possible because the solution for this geometry is "conical" and independent of "radial" distance in xyt-space, and this allows separation in appropriate coordinates. (This is Busemann's conical flow method [2] widely used in supersonic aerodynamics.) The method can also be used in problems for which the periodic solution is not known.

Before the conical flow method can be employed, it is necessary to know how the plane discontinuity surface propagates. The propagation of such discontinuities has been investigated by R. K. Luneberg [3] in electromagnetic theory, and by J. B. Keller [4] in acoustics. It is found in both cases that the discontinuity surface satisfies a first order partial differential equation, the eiconal equation in homogeneous media, and that the magnitude of the discontinuity varies in a simple manner as the surface moves. We make use of these results in the present investigation, and they enable us to convert the initial-boundary value problem into a characteristic-boundary value problem in xyt-space. The conical flow method is then used to obtain the solution.

The results apply to a single component of electric or magnetic field parallel to the edge of a perfectly conducting wedge or corner ($v = 0$ on wall corresponds

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

*This work was performed at Washington Square College of Arts and Science, New York University, and was supported in part by Contract No. AF-19(122)-42 with the U. S. Air Force through sponsorship of Geophysical Research Directorate, Air Force Cambridge Research Laboratories, Air Materiel Command.

to electric field, $\partial v/\partial n = 0$ to magnetic field). They also apply to acoustic pressure with rigid walls ($\partial v/\partial n = 0$) or free walls ($v = 0$).

In formulating the problem, we attempt to represent a plane pulse incident on a wedge or corner. However, for certain directions of incidence the pulse is in contact with the wedge at all times and thus a reflected pulse is always present. In the case of corners several reflected pulses may be present at all times. In these cases therefore we must include the reflected pulses in the formulation of the initial conditions.

In Section 2 we formulate the problem of a pulse incident on a wedge and state the initial conditions. In Section 3 we determine the subsequent behavior of the discontinuity surfaces. In Section 4 we introduce the method of conical flow and reduce the problem to the determination of a harmonic function in a circular sector. In Section 5 we solve for the harmonic function and thus complete the solution of the problem. In Section 6 we give the solutions for all cases of a pulse incident on a wedge. In Section 7 we treat the various cases of pulses incident in corners. In Section 8 we obtain the time-harmonic solution of the wedge problem as a Fourier integral of the pulse solution. In Section 9 the three dimensional case, in which the discontinuity surface at the pulse front is not parallel to the edge, is considered. Section 10 is the conclusion.

2. *Formulation*

We seek a solution of the wave equation.

$$v_{xx} + v_{yy} + v_{zz} - \frac{1}{c^2} v_{tt} = 0 \tag{1}$$

in the region $\varphi \leq \theta \leq 2\pi - \varphi$, where θ is the polar angle, $\theta = \arg(x + iy)$. The half-planes (walls) at $\theta = \pm\varphi$ form a wedge or corner according as φ is less or greater than $\pi/2$, the case $\varphi = \pi/2$ being trivial.

On the walls, we consider two types of boundary conditions:

Case A: $v = 0$; Case B: $\partial v/\partial n = 0$.

The solution which we consider will have jump discontinuities on certain moving surfaces, say $r(x, y, z) = ct$. We require that r satisfy the eiconal equation [2, 3]

$$r_x^2 + r_y^2 + r_z^2 = 1. \tag{2}$$

This implies that the surface can be constructed by Huygens' principle, that it moves with velocity c along its normal, and that it is reflected from the walls according to the law of reflection. We further assume that the reflected discontinuity is plus or minus the incident discontinuity according as the boundary condition is $\partial v/\partial n = 0$ or $v = 0$.

The orthogonal trajectories of a family of discontinuity surfaces $S(t)$ are straight lines called rays. The set of rays through a small closed curve on a

discontinuity surface $S(t_0)$ is called a tube. Denote by dS_0 the area on $S(t_0)$ which spans the tube and by $[v_0]$ the jump in v across $S(t_0)$. Let the corresponding quantities on $S(t)$ be dS and $[v]$. Then we require that the magnitude of the discontinuity $[v]$ vary inversely as $(dS)^{1/2}$, that is:

$$\lim_{dS_0 \to 0} \left(\frac{dS}{dS_0}\right)^{1/2} = \frac{[v_0]}{[v]}. \tag{3}$$

Equation 3 permits $[v]$ to be computed from $[v_0]$ on the same ray, once the discontinuity surfaces are known.

We now give the formulation of the problem for the wedge in the case for which no reflected pulse occurs initially and the boundary conditions A prevail.

Problem 1A: On the walls the boundary condition is $v = 0$. We assume that initially, $v = 0$ on one side of a plane and $v = 1$ on the other side, while $v_t = 0$ everywhere. The plane front of the pulse approaches a wedge and is parallel to its edge, (see Figure 1). Because of the special nature of these initial conditions we also assume that the plane discontinuity is moving toward the edge, in order to assure a unique solution.

It is convenient to introduce the normal to the discontinuity plane (positive in the direction of motion), and to call it the ray direction. The angle

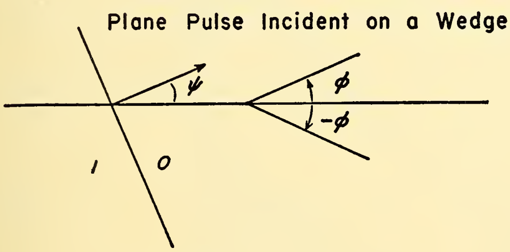


FIGURE 1

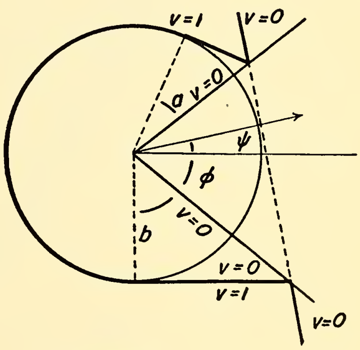


FIGURE 2

between the ray direction and the x-axis we call ψ, and assume it is positive. The problem formulated above makes sense only if $0 \leq \psi \leq \pi/2 - \varphi$. We will find it necessary later to distinguish the cases $0 \leq \psi \leq \varphi$ and $\varphi \leq \psi \leq \pi/2 - \varphi$.

3. *Propagation of Discontinuity*

It follows from equation (2) that a plane discontinuity surface moves parallel to itself with velocity c along its normal and from equation (3) that the

jump or discontinuity $[v] = 1$ across the pulse front does not change. This situation continues until the time the plane reaches the edge of the wedge when reflected and diffracted discontinuity surfaces may originate. These surfaces can be obtained by Huygen's principle (a consequence of equation 2) from the configuration at the instant of contact. One finds that the incident plane progresses parallel to itself and that one ($\psi \geq \varphi$) or two ($\psi < \varphi$) reflected plane discontinuity surfaces and a circular cylindrical surface with the edge as its axis are produced (see Figure 2). At all later times the configuration of surfaces is similar to that in Figure 2 and the scale is determined by a radius of the cylinder, which equals ct, if $t = 0$ is the instant of contact.

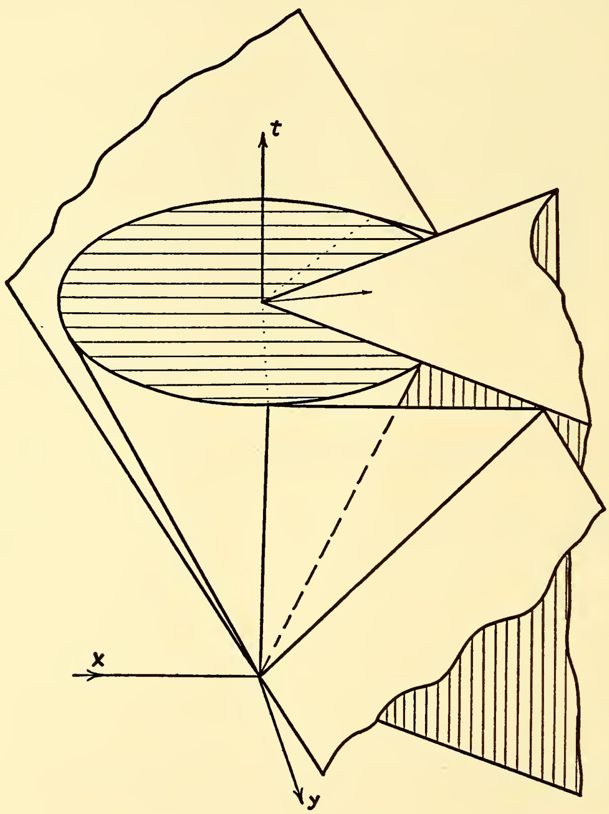


FIGURE 3

The jump across the original plane is unchanged, and the jump across the reflected plane (or planes) is its negative $[v] = -1$. The jump across the cylinder is zero, however, since all the rays reaching it come from the axis where $dS_0 = 0$ (see equation 3). Thus v is continuous across the cylinder. The value of v everywhere outside the cylinder is known (either 0 or 1) except in the one

($\psi \geq \varphi$) or two ($\psi < \varphi$) "triangular" regions bounded by the wedge, a reflected plane and the circle (see Figure 2). Since $v = 0$ on the wedge and behind the reflected plane, we assume $v = 0$ everywhere in this "triangle". This seems reasonable since the effect of the edge has not reached this region and therefore the reflection ought to be the same as from an infinite plane. We will validate the assumption by constructing a solution consistent with it.[1] The value of v is now known everywhere outside the circular sector bounded by the wedge and the circle. Since $v = 0$ on the wedge and v is continuous across the circular arc the values on the boundary are known. From these values we shall be able to determine v within the sector.

4. *Conical Flow Method*

Since the boundary data are independent of z, we seek a solution independent of z. Setting, in equation (1), $v_{zz} = 0$, we obtain

$$v_{xx} + v_{yy} - \frac{1}{c^2} v_{tt} = 0. \tag{4}$$

Let us consider the configuration of discontinuity surfaces in xyt-space. The circle of Figure 2 describes a characteristic cone, and the lines of that figure describe planes (see Figure 3). The solution v is constant in each of the regions outside the cone. Clearly the boundary values of v on the cone are constant along each generator. The boundary conditions are also constant on the wedge. Thus the boundary data are constant along radial lines through the origin, which we locate at the vertex of the cone. We therefore seek a solution v within the conical sector which is constant along each radial line.

In order to take advantage of the above assumption, we introduce special "polar" coordinates in xyt-space:

$$\begin{aligned} p &= [c^2t^2 - (x^2 + y^2)]^{1/2} \\ q &= \frac{ct}{p} \\ \theta &= \tan^{-1} y/x. \end{aligned} \tag{5}$$

This transformation is real within and on the cone. The surface of the cone is given by

$$p = 0, \qquad q = \infty. \tag{6}$$

In these coordinates equation (4) becomes

$$(p^2 v_p)_p + [(1 - q^2)v_q]_q + \frac{1}{1 - q^2} v_{\theta\theta} = 0. \tag{7}$$

[1]The statement about the field in the "triangular" regions need not be assumed but may be proved from the other assumptions and the stipulation of conical symmetry.

In accordance with our assumption that v is constant along radial lines, we set $v = v(q, \theta)$. Equation (7) now simplifies to

$$[(1 - q^2)v_q]_q + \frac{1}{1 - q^2} v_{\theta\theta} = 0. \tag{8}$$

Put

$$\rho = \left(\frac{q - 1}{q + 1}\right)^{1/2} \tag{9}$$

in equation (8). This yields Laplace's equation

$$\rho \frac{\partial}{\partial \rho}\left(\rho \frac{\partial v}{\partial \rho}\right) + \frac{\partial^2 v}{\partial \theta^2} = 0. \tag{10}$$

The solution of (10) may be written in the form

$$v = \mathscr{I}m\, f(z) \tag{11}$$

where $f(z)$ is an analytic function of $z = \rho e^{i\theta}$. Introducing $R = (x^2 + y^2)^{1/2}$ we have from equations (5) and (9)

$$z \equiv \rho e^{i\theta} = \frac{x + iy}{ct + (c^2t^2 - R^2)^{1/2}}; \qquad \rho = \frac{R}{ct + (c^2t^2 - R^2)^{1/2}}. \tag{12}$$

The cone $R \leq ct$ thus is mapped into the unit circle $\rho \leq 1$. The problem is now reduced to that of finding a function analytic in an appropriate sector of the unit circle with prescribed imaginary part on the boundary.

5. *Solution of the Problem*

The values of v on the boundary of the circular sector in Figure 2 are, in the case $\psi < \phi$:

$$v = 0 \quad \text{on} \quad \begin{cases} 0 \leq \rho \leq 1,\ \theta = \phi \\ \rho = 1,\ \phi \leq \theta \leq \phi + a \end{cases}$$

$$v = 1 \quad \text{on} \quad \rho = 1,\ \phi + a < \theta < 2\pi - \phi - b$$

$$v = 0 \quad \text{on} \quad \begin{cases} \rho = 1,\ 2\pi - \phi - b < \theta \leq 2\pi - \phi \\ 0 \leq \rho \leq 1,\ \theta - 2\pi - \phi. \end{cases}$$

Here $a = \phi - \psi$, $b = \phi + \psi$.

In order to solve for v, we map the exterior of the wedge in the z-plane into the upper half of the w-plane by the transformation:

$$w = re^{i\omega} = (e^{-i\phi}z)^{\lambda} \tag{13}$$

where $\lambda = \pi/(2\pi - 2\phi)$. Thus

$$r = \rho^{\lambda}, \qquad \omega = \lambda(\theta - \phi) = \lambda(\theta - \pi) + \pi/2. \tag{14}$$

The circular sector in which v is to be determined becomes a semicircle in the w-plane with $v = 0$ on the diameter (into which the sides of the wedge

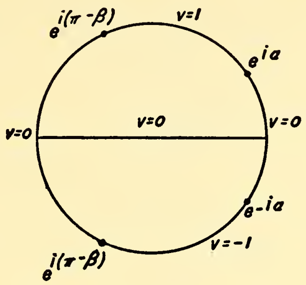


FIGURE 4

transform). By the reflection principle we may extend v into the whole circle, and obtain a boundary value problem in the unit circle. (See Figure 4; $\beta = \lambda b$, $\alpha = \lambda a$.)

In this case, as in all the others, we have to determine a harmonic function v with piecewise constant boundary values. The solution of the problem may be obtained as the sum of solutions which take on a specified constant value on one arc of the circle, the value zero on the rest. Let us write down the solution of this special problem once for all. Suppose $\omega_2 > \omega_1$ ($\omega_2 - \omega_1 < 2\pi$) and $v = c$ on the arc $\omega_2 \geq \omega \geq \omega_1$ and $v = 0$ elsewhere. It is not hard to show that v may be written in the form

$$v = \frac{c}{\pi}\left[\arg\left\{\frac{w - \exp\{i\omega_2\}}{w - \exp\{i\omega_1\}}\right\} - \frac{\omega_2 - \omega_1}{2}\right]. \tag{15a}$$

In terms of real variables we then have

$$v = \frac{c}{\pi}\arctan\left\{\frac{(1 - r^2)\sin\left(\frac{\omega_2 - \omega_1}{2}\right)}{(1 + r^2)\cos\frac{\omega_2 - \omega_1}{2} - 2r\cos\left(\omega - \frac{\omega_2 + \omega_1}{2}\right)}\right\}. \tag{15b}$$

The arctangent is taken in the interval between 0 and π. The solution to 1A may then be written explicitly,

$$\begin{aligned} v = {} & \frac{1}{\pi}\arctan\left\{\frac{-(1 - \rho^{2\lambda})\sin\lambda\pi}{2\rho^{\lambda}\cos\lambda(\theta + \psi - \pi) + (\rho^{2\lambda} + 1)\cos\lambda\pi}\right\} \\ & - \frac{1}{\pi}\arctan\left\{\frac{(1 - \rho^{2\lambda})\sin\lambda\pi}{2\rho^{\lambda}\cos\lambda(\theta - \psi - \pi) - (\rho^{2\lambda} + 1)\cos\lambda\pi}\right\}. \end{aligned} \tag{16}$$

6. *Pulse Incident on Wedge in General*

In order to specify the initial and boundary conditions properly we must distinguish several cases:

(1) $0 \leq \varphi \leq \pi/2$, $0 \leq \psi < \pi/2 - \varphi$ (*no initial reflected pulse*)

(a) $0 \leq \psi \leq \varphi$ (Figure 2)
(b) $\varphi \leq \psi \leq \pi/2 - \varphi$ (Figure 5)

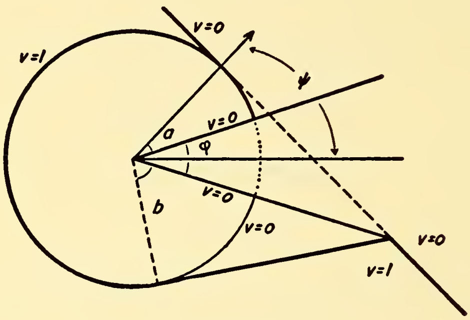


FIGURE 5

Both cases occur if $\varphi < \pi/4$; if $\varphi > \pi/4$ only case a occurs.

(2) $\pi/2 - \varphi \leq \psi < \pi - \varphi$ (*reflected pulse present initially*)

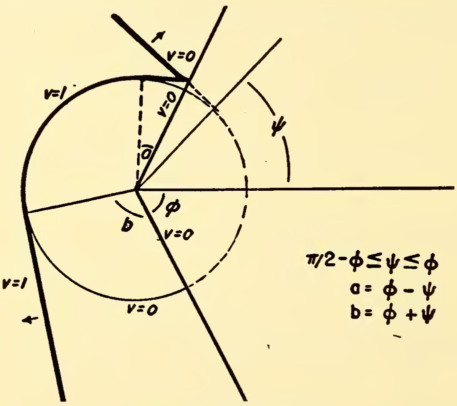


FIG. 6. Case 2a *A*.

(a) $\pi/2 - \varphi \leq \psi \leq \varphi$ (Figure 6)
The solution in the diffraction region (circular sector) has the same form as in case 1a above.

(b) $\varphi \leq \psi < \pi - \varphi$ (Figure 7)
The solution in the diffraction region (circular sector) has the same form as in case 1b above.

Both cases occur if $\varphi \geq \pi/4$; if $\varphi < \pi/4$ only case b occurs.

In each of these four cases we may consider either boundary condition A ($v = 0$) or B ($\partial v/\partial n = 0$) on the wedge. The wave fronts are shown in Figures 2, 5, 6 and 7. In the preceding section we obtained the solution (equation 16)

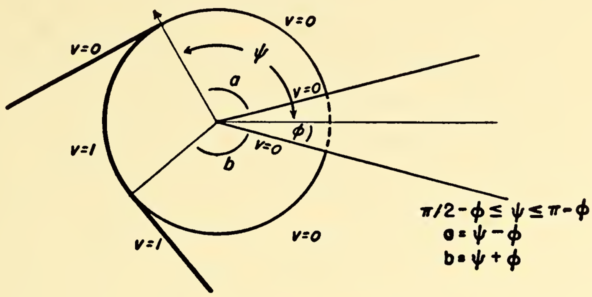


FIG. 7. Case 2b A.

for case 1a with boundary condition A (i.e., $0 \leq \psi \leq \varphi$, $v = 0$ on the wedge). In this section we shall give the solution for 1b with boundary condition A as well as for both problems 1a, 1b, with boundary condition B. This will complete the solution of all the wedge problems mentioned above, since the solutions of cases 2a and 2b are of the same form as those of cases 1a and 1b in the diffraction region. The solutions are obtained just as in the preceding section.

(1b A) $\varphi \leq \psi \leq \pi/2 - \varphi$; Figure 5; $v = 0$ on wedge.

$$v = \frac{1}{\pi}\arctan\left\{\frac{(1-\rho^{2\lambda})\cos\lambda\psi}{(1+\rho^{2\lambda})\sin\lambda\psi - 2\rho^{\lambda}\sin\lambda\theta}\right\} - \frac{1}{\pi}\arctan\left\{\frac{(1-\rho^{2\lambda})\cos\lambda\psi}{(1+\rho^{2\lambda})\sin\lambda\psi - 2\rho^{\lambda}\sin\lambda(\theta-2\pi)}\right\} \tag{17}$$

(1a B) $0 \leq \psi \leq \varphi$: Figure 2; $\partial v/\partial n = 0$ on wedge.

$$v = 1 + \frac{1}{\pi}\arctan\left\{\frac{(1-\rho^{2\lambda})\cos\lambda(\psi-\pi)}{(1+\rho^{2\lambda})\sin\lambda(\psi-\pi) - 2\rho^{\lambda}\sin\lambda(\theta-\pi)}\right\} + \frac{1}{\pi}\arctan\left\{\frac{-(1-\rho^{2\lambda})\cos\lambda(\psi+\pi)}{(1+\rho^{2\lambda})\sin\lambda(\psi+\pi) - 2\rho^{\lambda}\sin\lambda(\theta-\pi)}\right\} \tag{18}$$

(1b B) $\varphi \leq \psi \leq \pi/2 - \varphi$: Figure 5; $\partial v/\partial n = 0$ on wedge.

$$v = 1 - \frac{1}{\pi}\arctan\left\{\frac{-(1-\rho^{2\lambda})\cos\lambda(\psi-\pi)}{(1+\rho^{2\lambda})\sin\lambda(\psi-\pi) - 2\rho^{\lambda}\sin\lambda(\theta-\pi)}\right\} + \frac{1}{\pi}\arctan\left\{\frac{-(1-\rho^{2\lambda})\cos\lambda(\psi+\pi)}{(1+\rho^{2\lambda})\sin\lambda(\psi+\pi) - 2\rho^{\lambda}\sin\lambda(\theta-\pi)}\right\}. \tag{19}$$

7. *Pulses Incident in Corners*

We now consider the solution of the wave equation in the region $\pi - \varphi \leq \theta \leq \pi + \varphi$ where $\varphi \leq \pi/2$, and we call the region a corner. We assume all conditions of the problem, except the initial conditions, are the same as in the wedge problem. The initial conditions correspond to an incident plane pulse and ψ the direction of the incident ray, satisfies $\varphi > \psi \geq 0$.

In addition to the incident plane pulse, a number of reflected pulses will be present initially.[2] According to equation (2) the reflected discontinuity surfaces (or the rays orthogonal to them) satisfy the law of reflection. These plane discontinuity surfaces will, by Huygens' principle, move parallel to themselves into the corner with velocity c, be reflected from the walls, and ultimately emerge. Besides the reflected plane discontinuity surfaces, Huygens' principle yields an additional discontinuity surface bounding a region of diffraction consisting of a circular cylinder with the edge as axis. The solution will be piecewise constant everywhere except within the circular cylinder. We now consider the solution in this region.

First, we must consider the multiply reflected discontinuity planes. To this end, we investigate the reflections of the incident rays—lines normal to the incident discontinuity plane. An incident ray may strike either the upper $(\theta = \pi - \varphi)$ or lower wall $(\theta = \pi + \varphi)$ first. Accordingly we designate the ray as of type I or type II. These types include all rays except one which strikes the edge first, which can be disregarded for the purpose of determining the reflected discontinuity planes (see Figure 10).

The ray direction ψ_ν of a ray of type I after ν reflections is

$$\psi_\nu = (-1)^\nu(\psi + 2\nu\varphi). \tag{20}$$

This indicates that a reflection increases ψ by 2φ and changes its sense. For a ray of type II, the ray direction φ_ν after ν reflections is

$$\varphi_\nu = (-1)^{\nu+1}(2\nu\varphi - \psi). \tag{21}$$

The total number n of reflections suffered by a ray of type I is the unique integer for which

$$\pi - \varphi \leq \psi + 2n\varphi < \pi + \varphi. \tag{22}$$

Thus

$$\frac{\pi - \psi}{2\varphi} - \frac{1}{2} \leq n < \frac{\pi - \psi}{2\varphi} + \frac{1}{2}. \tag{23}$$

If n is odd the final reflection is from the upper wall $(\theta = \pi - \varphi)$; if n is even it is from the lower wall $(\theta = \pi + \varphi)$.

---

[2]See Figures 8 and 9.

Similarly the number m of reflections suffered by a ray of type II is given by

$$\frac{\pi + \psi}{2\varphi} - \frac{1}{2} \leq m < \frac{\pi + \psi}{2\varphi} + \frac{1}{2}. \tag{24}$$

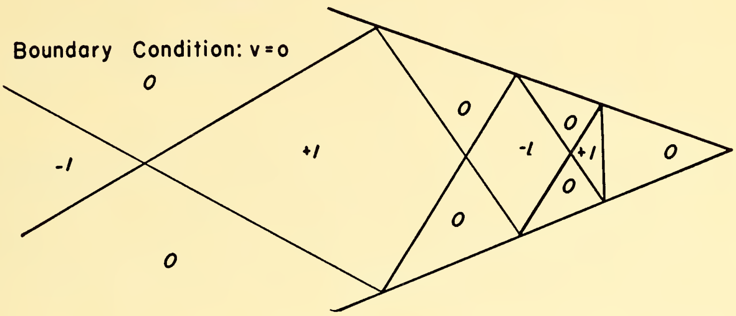


FIGURE 8

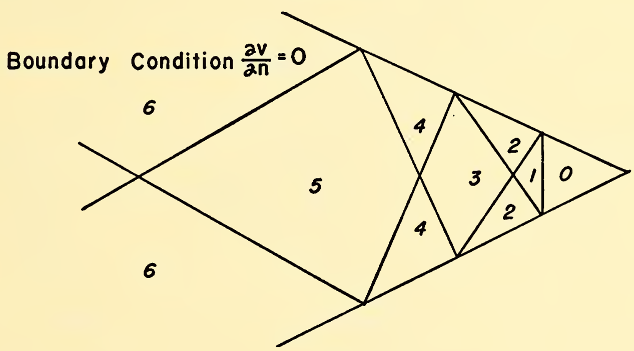


FIGURE 9

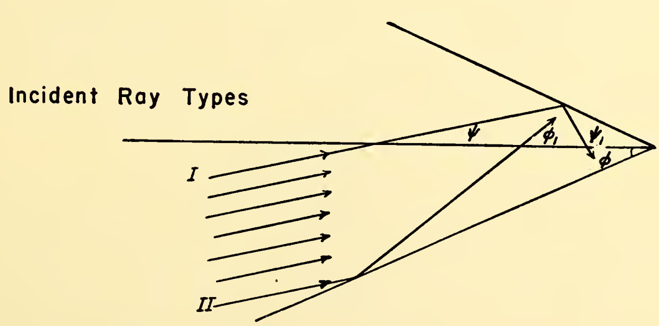


FIGURE 10

The last reflection is from the lower or upper wall according as m is odd or even.

These results may be written explicitly in terms of the number theoretic

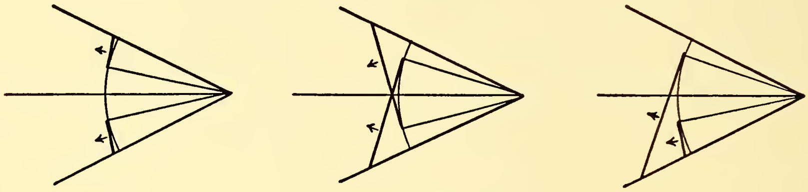
function, $[\lambda]$, the largest integer in λ. If we set $\lambda = (\pi/2\varphi)$, $k = \lambda - [\lambda]$ we obtain

$$n = \begin{cases} [\lambda] + 1, & \text{for} \quad \psi < \varphi(2k - 1) \\ [\lambda], & \text{for} \quad \psi \geq \varphi(2k - 1) \end{cases}$$

$$m = \begin{cases} [\lambda] + 1, & \text{for} \quad \psi > \varphi(1 - 2k) \\ [\lambda], & \text{for} \quad \psi \leq \varphi(1 - 2k). \end{cases}$$

From a knowledge of the ultimate ray directions, given by equations (20) to (24), the ultimate reflected plane discontinuity surfaces may be determined since they are orthogonal to the rays. There are four kinds of configurations which may occur with regard to the ultimate wave fronts (i.e. planes normal to ultimate rays):

Case a: m,n have the same parity and the wave fronts do not overlap.
Case b: m,n have the same parity and the wave fronts overlap.
Case c: m,n have opposite parity, m *odd*, n *even*.
Case d: n odd, m even is symmetric to c.



Case a) Case b) Case c)

In addition to the ultimate plane wave fronts, Huygens' principle yields a circular cylindrical wave front with the edge as axis and tangent to the ultimate wave fronts at their edges. As we stated above, the solution is piecewise con-

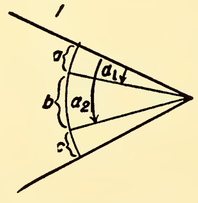


FIGURE 11

stant everywhere except within this circle which thus contains all the diffraction effects. In order to find the solution within this circle we are led by the preceding considerations to a boundary value problem where the boundary values are constant on each of three segments of the arc (see Figure 11). In solving

the boundary value problems we shall use the notation a, b, c, for the boundary values and α_1, α_2, for the angles, as indicated in the diagram.

In all these problems we map the corner into a semicircle by means of the transformation

$$w = r\,e^{i\omega} = (z \exp \{i(\varphi - \pi)\})^{\lambda} \tag{25}$$

where

$$\lambda = \frac{\pi}{2\varphi}. \tag{26}$$

In this mapping $r = \rho^{\lambda}$, $\omega = \lambda(\theta + \varphi - \pi) = \lambda(\theta - \pi) + \pi/2$. In particular, we set

$$\omega_1 = \lambda\alpha_1 \qquad \omega_2 = \lambda\alpha_2 . \tag{27}$$

As we have seen, the angles ω_1, ω_2 may be completely specified. To state the problem completely it is only necessary to find the boundary values. There are two cases.

A. $v = 0$ *on the Walls*

The field strength is assumed to be 1 behind the initial wave front and to be zero ahead. The first reflected wave fronts will each add -1 to the field already present, the second wave fronts, $+1$, then -1, $+1$, $\cdots$ in alternation. Our computation of the number of reflections then tells us what the field is in the region ahead of *both* the final wave fronts, namely

$$\sum_{\nu=0}^{n-1} (-1)^{\nu} + \sum_{\nu=0}^{m-1} (-1)^{\nu} - 1 = \frac{(-1)^{n+1} + (-1)^{m+1}}{2}. \tag{28}$$

To compute the boundary conditions we simply note that if m (say) is odd the wave front adds -1 to the field as it passes over, if m is even it adds $+1$ (see Figure 8).

B. $(\partial v/\partial n) = 0$ *on the Walls*

In this case each reflected wave front adds $+1$ to the already present field. The region ahead of both final wave fronts therefore has the field strength

$$n + m - 1. \tag{29}$$

Each of the two last wave fronts adds 1 to this value (see Figure 9).

The accompanying table gives all possibilities.[3]

In each of the problems we are led by the reflection principle to a boundary value problem on the unit circle. The circle may be divided into four arcs on each of which the boundary values are constant. We note in each case that two of the constant values are equal. The problem may then be somewhat simplified by subtracting this value from the boundary values.

---

[3]Letters in parentheses refer to the cases of page 86(S150).

Problem	Parity n	Parity m	Case	Relation between ψ_n and φ_m	α_1	α_2	Boundary Values a	Boundary Values b	Boundary Values c
A1	odd	odd	(a)	$\varphi_m - \psi_n \geq 2\pi$	$\psi_n+\varphi+\pi$	$\varphi_m+\varphi-\pi$	0	1	0
B1							$n+m$	$n+m-1$	$n+m$
A1′			(b)	$\varphi_m - \psi_n \leq 2\pi$	$\varphi_m+\varphi-\pi$	$\psi_n+\varphi+\pi$	0	-1	0
B1′							$n+m$	$n+m+1$	$n+m$
A2	even	even	(a)	$\psi_n - \varphi_m \geq 2\pi$	$\varphi_m+\varphi+\pi$	$\psi_n+\varphi-\pi$	0	-1	0
B2							$n+m$	$n+m-1$	$n+m$
A2′			(b)	$\psi_n - \varphi_m \leq 2\pi$	$\psi_n+\varphi-\pi$	$\varphi_m+\varphi+\pi$	0	1	0
B2′							$n+m$	$n+m+1$	$n+m$
A3	even	odd	(c)	$\psi_n \leq \varphi_m$	$\psi_n+\varphi-\pi$	$\varphi_m+\varphi-\pi$	0	1	0
B3							$n+m+1$	$n+m$	$n+m+1$
A3′			(c)	$\psi_n \geq \varphi_m$	$\varphi_m+\varphi-\pi$	$\psi_n+\varphi-\pi$	0	-1	0
B3′							$n+m+1$	$n+m$	$n+m+1$
A4	odd	even	(d)	$\psi_n \leq \varphi_m$	$\psi_n+\varphi+\pi$	$\varphi_m+\varphi+\pi$	0	-1	0
B4							$n+m+1$	$n+m$	$n+m-1$
A4′			(d)	$\psi_n \geq \varphi_m$	$\varphi_m+\varphi+\pi$	$\psi_n+\varphi+\pi$	0	1	0
B4′							$n+m+1$	$n+m$	$n+m-1$

Let us now construct the solutions. For all the problems A set (in view of Equation (15b):

$$\sigma = \frac{1}{\pi}\arctan\left\{\frac{(1-r^2)\sin\left(\dfrac{\omega_2-\omega_1}{2}\right)}{(1+r^2)\cos\left(\dfrac{\omega_2-\omega_1}{2}\right) - 2r\cos\left(\omega - \dfrac{\omega_2+\omega_1}{2}\right)}\right\}$$

(30)

$$\tau = \frac{1}{\pi}\arctan\left\{\frac{(1-r^2)\sin\left(\dfrac{\omega_2-\omega_1}{2}\right)}{(1+r^2)\cos\left(\dfrac{\omega_2-\omega_1}{2}\right) - 2r\cos\left(\omega + \dfrac{\omega_2+\omega_1}{2}\right)}\right\}$$

where $\omega_1 = \lambda\alpha_1$ and $\omega_2 = \lambda\alpha_2$.

Here the values of the arctangents are restricted to the interval between 0 and π. For A1, A2′, A3, A4′ we have $v = \sigma - \tau$. For A1′, A2, A3′, A4, $v = \tau - \sigma$.

In the remaining problems, B1, B1′, B2, B2′, B3, B3′, B4, B4′ set

$$
\begin{aligned}
\xi &= \frac{1}{\pi} \arctan \left\{ \frac{(1 - r^2) \sin \omega_1}{(1 + r^2) \cos \omega_1 - 2r \cos \omega} \right\} \\
\eta &= \frac{1}{\pi} \arctan \left\{ \frac{-(1 - r^2) \sin \omega_2}{(1 + r^2) \cos \omega_2 - 2r \cos \omega} \right\}
\end{aligned} \tag{31}
$$

where the values of the arctangents are taken between 0 and π.

For B1, B2 we have $v = n + m - 1 + \xi + \eta$. For B1′, B2′ we have $v = n + m + 1 - \xi - \eta$. For B3, B3′ we have $v = n + m - \xi + \eta$. For B4, B4′ we have $v = n + m + \xi - \eta$.

The solutions may easily be calculated by taking the values of α_1 and α_2 from the tables. Using

$$
\begin{aligned}
\psi_n &= (2n\varphi + \psi)(-1)^n \\
\varphi_m &= (2m\varphi - \psi)(-1)^{m+1}
\end{aligned} \tag{32}
$$

we find as the solution to the problems A1 A1′, A2, A2′

$$
\begin{aligned}
v = &-\frac{1}{\pi} \arctan \left\{ \frac{(1 - \rho^{2\lambda}) \sin \lambda\pi}{(1 + \rho^{2\lambda}) \cos \lambda\pi + 2\rho^{\lambda} \cos \lambda(\theta + \psi - \pi)} \right\} \\
&+ \frac{1}{\pi} \arctan \left\{ \frac{(1 - \rho^{2\lambda}) \sin \lambda\pi}{(1 + \rho^{2\lambda}) \cos \lambda\pi - 2\rho^{\lambda} \cos \lambda(\theta - \psi - \pi)} \right\}
\end{aligned} \tag{33}
$$

as solutions to the problems A3, A3′, A4, A4′

$$
\begin{aligned}
v = &\frac{1}{\pi} \arctan \left\{ \frac{(1 - \rho^{2\lambda}) \cos \lambda\psi}{(1 + \rho^{2\lambda}) \sin \lambda\psi - 2\rho^{\lambda} \sin \lambda\theta} \right\} \\
&- \frac{1}{\pi} \arctan \left\{ \frac{(1 - \rho^{2\lambda}) \cos \lambda\psi}{(1 + \rho^{2\lambda}) \sin \lambda\psi - 2\rho^{\lambda} \sin \lambda(\theta - 2\pi)} \right\}
\end{aligned} \tag{34}
$$

and as solutions to all the problems B

$$
\begin{aligned}
v = n + m &+ \frac{1}{\pi} \arctan \left\{ \frac{(1 - \rho^{2\lambda}) \cos \lambda(\psi - \pi)}{(1 + \rho^{2\lambda}) \sin \lambda(\psi - \pi) - 2\rho^{\lambda} \sin \lambda(\theta - \pi)} \right\} \\
&- \frac{1}{\pi} \arctan \left\{ \frac{(1 - \rho^{2\lambda}) \cos \lambda(\psi + \pi)}{(1 + \rho^{2\lambda}) \sin \lambda(\psi + \pi) - 2\rho^{\lambda} \sin \lambda(\theta - \pi)} \right\}.
\end{aligned} \tag{35}
$$

In connecting all the problems which have the same formal solution we have used the identity

$$\arctan(-x) = \pi - \arctan x \tag{36}$$

valid when the arctangents are restricted to the interval between 0 and π.

We observe that the above solutions apply also to the wedge problem if we permit φ to range through the entire interval $0 < \varphi \leq \pi$. In fact if we replace φ by $\pi - \varphi$ in these equations we obtain the solutions to the wedge problem given previously.

There remains the interesting question as to when there is no diffraction phenomenon, i.e, v is constant within the circle. This will occur under the following conditions for either boundary condition

$$\left.\begin{array}{lll} \alpha) & m, n \text{ odd}; & \varphi_m - \psi_n = 2\pi \\ \beta) & m, n \text{ even}; & \psi_n - \varphi_m = 2\pi \end{array}\right\} \quad \varphi = \frac{\pi}{m+n} \quad \text{independent of } \psi.$$

When λ is an odd integer the emergent ray lies in the direction $\pi - \psi$. This "mirror" property is the ordinary reflection principle for the plane, $\varphi = \pi/2$. When λ is an even integer, the emergent ray lies in a direction opposite to that of the incident ray. This is the familiar property of the right-angled corner, $\varphi = \pi/4$.

8. *The Time-Harmonic Solution*

By employing Duhamel's theorem, the solution of the above problems for an incident plane periodic wave, or wave of any other time dependence, may be obtained. Thus MacDonald's solution of the wedge problem or Sommerfeld's solution of the half-plane ($\varphi = 0$) problem may be obtained.

We consider the wedge problem 1Aa. Duhamel's theorem states that the time periodic solution $E = v(r, \theta, \omega) \exp\{-i\omega t\}$ may be derived from the pulse solution $u(r, \theta, t)$ according to the formula

$$v(\omega) = -i\omega \int_{-\infty}^{\infty} u(t) \exp\{i\omega t\}\, dt. \tag{37}$$

Reciprocally, the pulse solution is given by

$$u(t) = -\frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{v(\omega)}{\omega} \exp\{-i\omega t\}\, d\omega. \tag{38}$$

Knowing the function u, we have an expression for v as a Fourier integral.

First, we write the representation of u in the region $ct > R$ by means of the formula (15a),

$$u = \frac{1}{\pi}\left\{\arg\frac{1 - r\exp\{i(\omega - \pi + \beta)\}}{1 - r\exp\{i(\omega - \alpha)\}} - \arg\frac{1 - r\exp\{i(\omega + \alpha)\}}{1 - r\exp\{i(\omega - \pi - \beta)\}}\right\}.$$

Here u is the solution of problem 1Aa and it will be recalled that

$$r = \left[\frac{R}{ct + (c^2t^2 - R^2)^{1/2}}\right]^{\lambda}, \qquad \alpha = \lambda(\varphi - \psi), \qquad \beta = \lambda(\varphi + \psi), \tag{39}$$

$$\omega = \lambda(\theta - \varphi), \qquad \lambda\varphi = \lambda\pi - \frac{\pi}{2}.$$

From the series representation

$$\arg(1 - re^{i\chi}) = -\sum_{n=1}^{\infty}\frac{r^n}{n}\sin n\chi$$

we have

$$u = \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{r^n}{n}A_n$$

where

$$\begin{aligned} A_n &= \sin n(\omega + \alpha) + \sin n(\omega - \alpha) \\ &\quad - \sin n(\omega - \pi + \beta) - \sin n(\omega - \pi - \beta) \\ &= 4\sin n\lambda(\theta - \varphi)\sin n\lambda(\psi - \varphi + \pi)\sin n\lambda\pi. \end{aligned} \tag{40}$$

We seek a function v_n satisfying

$$r^n = \left[\frac{R}{ct + (c^2t^2 - R^2)^{1/2}}\right]^{n\lambda} = -\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{v_n(\omega)}{\omega}\exp\{-i\omega t\}\,dt$$

for $ct > R$.

From Campbell and Foster [5] (909.7) we have for $a, \nu > 0$

$$-\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{J_\nu(a\omega)}{\omega}\exp\{-i\omega t\}\,d\omega$$

$$= \begin{cases} \dfrac{\exp\{i\nu\pi/2\}}{\pi\nu}\sin\left[\nu\cos^{-1}\left(-\dfrac{t}{a}\right)\right] & 0 \le t \le a \\ \dfrac{\exp\{i\nu\pi/2\}\sin\pi\nu}{\pi\nu}\left[\dfrac{t}{a} + \left(\dfrac{t^2}{a^2} - 1\right)^{1/2}\right]^{-\nu} & a \le t. \end{cases} \tag{41}$$

This leads us to take

$$v_n = \frac{n\lambda\pi}{\sin n\lambda\pi}\exp\{-in\lambda\pi/2\}J_{n\lambda}(kR)$$

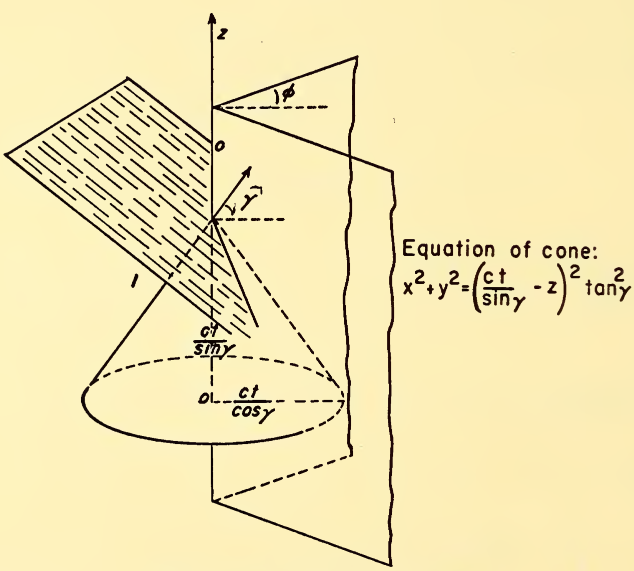
where $k = \omega/c$. With this formula the full periodic solution ought to be

$$(42) \qquad v = 4\lambda \sum_{n=1}^{\infty} \exp\{-in\lambda\pi/2\} J_{n\lambda}(kR) \sin n\lambda(\theta - \varphi) \sin n\lambda(\psi - \varphi + \pi).$$

This agrees with MacDonald's solution. It is in fact the correct solution for our problem, since a relatively simple computation shows that the function u in the region $0 \leq ct \leq R$ is actually given by (41).

9. *Three Dimensional Case*

If the incident discontinuity surface or pulse front is not parallel to the edge of the wedge or corner, it will intersect the edge at all times. Thus reflected and diffracted discontinuity surfaces will be present at all times, and must be included in giving the initial conditions. If these surfaces and the field are given correctly at the initial instant, we expect these surfaces as well as the field distribution to remain geometrically congruent at all times. This, then, is the condition on the initial conditions.



Plane Pulse at Skew Incidence

FIGURE 12

We assume that the discontinuity surfaces consist of an incident plane, two reflected plane sections (for a pulse incident on a wedge) and a diffracted cone (see Figure 12). We denote by γ the angle between the incident discontinuity plane and the edge. It is easily seen, as in Section 3, that this configuration of

surfaces persists with the point of intersection moving along the edge with velocity $c/\sin\gamma$. The discontinuities across these surfaces are also determined as in Section 3 for scalar quantities (e.g. acoustic pressure). For electromagnetic problems the reflected discontinuities may be obtained by employing the electromagnetic boundary conditions (see [3]). In any case, the discontinuity will be zero across the cone and constant across the planes.

To solve for v, we introduce the new (moving) coordinate

$$\zeta = z - (ct/\sin\gamma). \tag{43}$$

In the x,y,ζ-coordinates the discontinuity surfaces are stationary, and equation (1) becomes

$$v_{xx} + v_{yy} - \frac{1}{\tan^2\gamma} v_{\zeta\zeta} = 0. \tag{44}$$

This equation is the same as equation 4 with t replaced by $\zeta \tan\gamma$. The boundary surfaces in Figure 12 are similar to those in Figure 3, and the boundary values are also constant along rays. Thus the conical flow method and solution described previously also apply in this case.

10. *Conclusion*

By combining the results of Luneberg on the propagation of electromagnetic discontinuities with Busemann's conical flow method for solving the wave equation, it has been possible to obtain exact, explicit, solutions of the two and three dimensional diffraction of pulses by wedges and corners. The results, in closed form, involve only elementary functions and also apply to acoustic problems. Since the diffracting surface must be both a cylinder and a cone in xyt-space, the only other surface which can be treated in this way is a wire of zero thickness and infinite length. However, the results do describe parts of the field resulting from diffraction of a pulse by any polyhedral surface.

It also seems likely that the method can be extended to treat diffraction of a pulse by a cylinder of polygonal cross section, as well as diffraction of a pulse by a number of parallel wires of zero thickness (grating). In both of these cases, the new difficulty arises from diffraction of a diffracted wave. It seems that a modification of Busemann's [2] infinitesimal conical flow will suffice for the solution of these problems, and this investigation is already in progress.

The major difficulty in determining the diffraction of a pulse from a polyhedron is the determination of the field within the spherical discontinuity surface arising from a vertex as is already apparent in the case of a trihedral angle.

By employing Duhamel's theorem, the solutions of the wedge and corner problems for an incident wave of arbitrary time dependence may be obtained. In particular, we have exhibited MacDonald's solution of the wedge problem for periodic time dependence. Sommerfeld's solution for the half-plane ($\phi = 0$) also may be obtained in this manner.

Bibliography

1. MacDonald, H. M., *Electromagneticism*, G. Bell & Sons, Ltd., London, 1934, p. 79 ff.
2. Busemann, A., *Infinitesimal conical supersonic flow*, Schriften der Deutschen Akademie für Luftfahrforschung, Volume 7B, No. 3, 1943, pp. 105–122.
3. Luneberg, R. K., *Mathematical Theory of Optics*, Brown University Lectures, 1944.
4. Keller, J. B., *Mechanics of Continuous Media*, New York University Lectures, 1949–1950.
5. Campbell, G. A., and Foster, R. M., *Fourier Integrals for Practical Applications*, Van Nostrand, New York, 1948.

Vector Wave Functions

By R. D. SPENCE AND C. P. WELLS
Michigan State College

1. *Introduction*

This paper reviews the old and rather familiar problem of finding solenoidal solutions of the vector wave equation. Since this problem has been rather extensively treated for the case of spherical and cylindrical coordinates by Hansen [1] and others we shall confine our attention primarily to the spheroidal systems.

The solutions of the scalar wave equation in these coordinates are now fairly well known.[1] Tables[2] of both the prolate and oblate wave functions are available, which, though somewhat limited in extent and argument interval, are sufficiently extensive to make it possible to solve many interesting problems. Even so the scalar functions are rather complicated compared to those one meets in connection with the analogous problem in spherical or cylindrical coordinates. Much of the difficulty arises from the fact that the angular and radial spheroidal wave functions depend on the frequency as a parameter as well as on the usual spatial variables. As a consequence simple recursion formulas and relations between functions and derivatives of various orders do not exist.

2. *Solenoidal Solutions of the Vector Wave Equation*

We now consider the problem of constructing a set of vector wave functions from scalar wave functions. To assure the solenoidal nature of our solutions we may write the desired vector $\mathbf{V}_n$ in the form

$$\mathbf{V}_n = \nabla \times (\mathbf{f}\psi_n) \tag{1}$$

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

[1]See for example Stratton, Morse, Chu, and Hutner, *Elliptic Cylinder and Spheroidal Wave Functions*, New York, John Wiley and Sons (1941). See also Leitner, A., and Spence, R. D., *The oblate spheroidal wave functions*, Journal of the Franklin Institute, Volume 249, 1950, p. 299.

[2]Tables of the oblate functions are given in the paper by Leitner and Spence. One of the present authors (R.D.S.) has available tables of the prolate functions.

where ψ_n is a solution of

$$\nabla^2 \psi_n + k^2 \psi_n = 0$$

in spheroidal coordinates. If one now inserts the assumed expression for $\mathbf{V}_n$ in the vector wave equation one finds that the unknown function $\mathbf{f}$ must satisfy

$$\nabla^2 \mathbf{f} + \nabla(\ln \psi_n^2) \cdot \nabla \mathbf{f} = \frac{\nabla \varphi_n}{\psi_n} \tag{2}$$

where the φ_n are a set of arbitrary scalar functions. If $\mathbf{f}$ is required to satisfy the conditions (a) $\mathbf{f}$ is independent of ψ_n , (b) $\nabla^2 \mathbf{f} = 0$, (c) $\mathbf{f}$ finite except at infinity, one finds

$$\mathbf{f} = K_1 \mathbf{a} \qquad \text{or} \qquad \mathbf{f} = K_2 \mathbf{r} \tag{3}$$

where $\mathbf{a}$ and $\mathbf{r}$ represent a constant vector and the position vector respectively and K_1 and K_2 are constants. The obvious cylindrical and spherical symmetry of the two solutions for $\mathbf{f}$ are responsible for the remarkable simplicity of the vector wave functions in cylindrical and spherical coordinates. A second solution normal to $\mathbf{V}_n$ is clearly

$$\mathbf{U}_n = \nabla \times \mathbf{V}_n . \tag{4}$$

3. *Orthogonality*

We now examine the set of functions $\{\mathbf{V}_n\}$ and $\{\mathbf{U}_n\}$ for those properties which may be of use in solving electromagnetic problems. In cylindrical and spherical coordinates one can easily show that vector wave functions are orthogonal, that is

$$\begin{aligned} &\int_s \mathbf{V}_n \cdot \mathbf{V}_m \, d\sigma = 0, \qquad n \neq m \\ &\int_s \mathbf{U}_n \cdot \mathbf{U}_m \, d\sigma = 0, \qquad n \neq m \\ &\int_s \mathbf{V}_n \cdot \mathbf{U}_n \, d\sigma = 0 \end{aligned} \tag{5}$$

where s is an appropriate surface, provided we take $\mathbf{f} = \mathbf{k}$ (a unit vector along the z-axis) for the cylindrical case and $\mathbf{f} = \mathbf{r}$ in the spherical case. In general we have

$$\int_s \mathbf{V}_n \cdot \mathbf{V}_m \, d\sigma = - \int_s (\nabla \psi_n \cdot \nabla \psi_m - \nabla \psi_n \cdot \mathbf{ff} \cdot \nabla \psi_m) \, d\sigma. \tag{6}$$

Let us now consider the result of making a "wrong" choice of $\mathbf{f}$ for a given set of ψ_n . For example we might choose $\mathbf{f} = \mathbf{k}$ and ψ_n the set of spherical wave

functions. The first integral on the right in (6) vanishes and we are left with

$$\int_s \mathbf{V}_n \cdot \mathbf{V}_m \, d\sigma = \int_s \frac{\partial \psi_n}{\partial z} \frac{\partial \psi_m}{\partial z} \, d\sigma \tag{7}$$

which does not vanish when integrated over the surface of a sphere. Thus we see that in general our wave functions are not orthogonal among themselves. Furthermore there is no reason to suppose that the two sets of functions $\{\mathbf{V}_n\}$ and $\{\mathbf{U}_n\}$ are orthogonal. For example if we again choose $\mathbf{f} = \mathbf{k}$ and let $\{\psi_n\}$ be either a set of spherical or spheroidal wave functions we can easily show that it is possible to expand certain functions in either the set $\{\mathbf{V}_n\}$ or $\{\mathbf{U}_n\}$. Thus

$$\begin{aligned} \mathbf{i}\, e^{ikz} &= \sum_n A_n \mathbf{V}_n = \sum_n A_n \nabla \times \mathbf{k}\psi_n \\ \mathbf{i}\, e^{ikz} &= \sum_n B_n \mathbf{U}_n = \sum_n B_n \nabla \times \nabla \times \mathbf{k}\psi_n\,. \end{aligned} \tag{8}$$

The coefficients A_n and B_n can easily be found by using the orthogonality of the scalar wave functions

$$A_n = -\frac{(2n+1)(-1)^n}{4ik}, \qquad B_n = \frac{1}{ik} A_n \tag{9}$$

for spherical coordinates.

The types of solenoidal vector wave functions for which pairs of expansions such as indicated in (8) exist can easily be found. Let

$$\mathbf{F} = \nabla \times \mathbf{k}\Psi = \nabla \times \nabla \times \mathbf{k}\Phi \tag{10}$$

where Ψ and Φ are solutions of the scalar wave equation such that

$$\begin{aligned} \Psi &= \sum_n a_n \psi_n \\ \Phi &= \sum_n b_n \psi_n \end{aligned} \tag{11}$$

where

$$\begin{aligned} \nabla^2 \psi_n + k^2 \psi_n &= 0 \\ \nabla^2 \Psi &= \sum_n a_n \nabla^2 \psi_n \\ \nabla^2 \Phi &= \sum_n b_n \nabla^2 \psi_n \end{aligned} \tag{12}$$

then

$$\begin{aligned} \Psi &= f_1(x, y) e^{ikz} + f_2(x, y) e^{-ikz} \\ \Phi &= g_1(x, y) e^{ikz} + g_2(x, y) e^{-ikz} \end{aligned} \tag{13}$$

where f_1, f_2, g_1 and g_2 are solutions of the two dimensional Laplace equation. The functions of f and g are related by

$$\frac{\partial f}{\partial y} = \frac{\partial^2 g}{\partial x\, \partial z}, \qquad \frac{\partial f}{\partial x} = -\frac{\partial^2 g}{\partial y\, \partial z}. \tag{14}$$

Let us now consider certain conditions under which a set of vector wave functions is orthogonal. If the vector wave functions are tangent to one of the sets of the coordinate surfaces, sufficient conditions for the orthogonality of the vector wave functions can easily be written down. Let u_1, u_2, u_3 represent the coordinates, $\mathbf{i}_1$, $\mathbf{i}_2$, $\mathbf{i}_3$ the corresponding unit vectors and h_1, h_2, h_3 the corresponding metrical coefficients. The set of vector wave functions $\mathbf{V}_n$ tangent to the surface u_1 = constant will be orthogonal on this surface provided

(15)

$$\text{(a)} \quad \mathbf{V}_n = \mathbf{Q}_n(u_2, u_3) f_n(u_1)$$

$$\text{(b)} \quad \int_s (V_{n3} V_{m2} - V_{n3} V_{m2}) \left(\mathbf{i}_2 \cdot \frac{\partial \mathbf{i}_3}{\partial u_1} - \mathbf{i}_3 \cdot \frac{\partial \mathbf{i}_2}{\partial u_1} \right) \frac{d\sigma}{h_1} = 0$$

where V_{n2} and V_{n3} are the components of $\mathbf{V}_n$ along u_2 and u_3 and V_{m2} and V_{m3} are the corresponding components of $\mathbf{V}_m$.

These equations are satisfied by the usual cylindrical and spherical vector wave functions $\nabla \times \mathbf{k}\psi_n^{\text{cyl.}}$ and $\nabla \times \mathbf{r}\psi_n^{\text{sph.}}$ respectively. We shall return to the problem of finding tangential solutions later on.

4. *Spheroidal Functions*

After the preliminary discussion it seems clear that it would be rather surprising to find that either choice of $\mathbf{f}$ yielded a set of vector wave functions orthogonal over the surface of a spheroid. Simple computations show that no orthogonality of the type (5) exists.

To show this we make use of the fact that two solutions $\mathbf{V}_n$ and $\mathbf{V}_m$ of the wave equation must satisfy

$$\int_s \left(\mathbf{V}_n \cdot \frac{\partial \mathbf{V}_m}{\partial u_1} - \mathbf{V}_m \cdot \frac{\partial \mathbf{V}_n}{\partial u_1} \right) \frac{d\sigma}{h_1} = 0 \tag{16}$$

integrated over the closed surface u_1 = constant. We assume that

$$\mathbf{V}_n = \nabla \times \mathbf{k}\psi_n \qquad \text{or} \qquad \nabla \times \mathbf{r}\psi_n \tag{17}$$

and that the ψ_n are separable solutions of the scalar wave equation in either oblate or prolate spheroidal coordinates. Then ψ_n has the form $F_n(u_1)G_n(u_2, u_3)$. It may be shown that (16) is incompatible with

$$\int_s \mathbf{V}_n \cdot \mathbf{V}_m \, d\sigma = 0.$$

This immediately poses the question as to just how one may hope to solve

a problem in spheroidal coordinates. The only help which seems immediate lies in the orthogonality of the scalar wave functions from which the vector wave functions are derived. In certain cases this appears to be sufficient. For example suppose we wish to expand

$$\mathbf{i}e^{-ikz}$$

in the functions $\mathbf{V}_n = \nabla \times \mathbf{r}\psi_n$ and $U_n = \nabla \times \nabla \times \mathbf{r}\psi_n$ where the ψ_n are oblate spheroidal wave functions. Since the function being expanded contains a radial component and since only the set $\{\mathbf{U}_n\}$ can supply this radial component it is clear that in this case we shall not find two independent expansions as in the case previously mentioned. On equating components and making use of the orthogonality of the scalar functions one finds

$$\mathbf{i}e^{-ikz} = \sum_l C_l(\mathbf{V}_{l1_o} + i\mathbf{U}_{l1_e}) \tag{18}$$

where

$$\begin{aligned} \mathbf{V}_{l1_o} &= \nabla \times \mathbf{r}(u_{l1}(\eta)\,{}^{(1)}\mathbf{v}_{l1}(\xi) \sin \varphi) \\ \mathbf{U}_{l1_e} &= \frac{1}{k} \nabla \times \nabla \times \mathbf{r}(u_{l1}(\eta)\,{}^{(1)}v_{l1}(\xi) \cos \varphi), \end{aligned} \tag{19}$$

$$\begin{aligned} C_l &= \frac{-i}{N_{l1}^{(1)}v_{l1}(0)ka} \int_{-1}^{+1} \frac{1 - \cos ka(1-\eta^2)^{1/2}}{(1-\eta^2)^{1/2}} u_{l1}(\eta)\, d\eta \cdots, \qquad l \text{ odd} \\ C_l &= \frac{1}{N_{l1}^{(1)}v_{l1}(0)ka} \int_{-1}^{+1} \frac{\eta[\sin ka(1-\eta^2)^{1/2} - ka(1-\eta^2)^{1/2}]}{1-\eta^2} u_{l1}(\eta)\, d\eta \cdots, \qquad l \text{ even} \end{aligned} \tag{20}$$

$$N_{l1} = \int_{-1}^{+1} [u_{l1}(\eta)]^2\, d\eta$$

In the above $u_{lm}(\eta)$ represents the "angular" oblate spheroidal wave function, ${}^{(1)}v_{lm}(\xi)$ the first kind radial oblate spheroidal wave function, $k = 2\pi/\lambda$ and a is the radius of the focal circle.

The availability of a plane wave expansion such as we have previously indicated leads one to consider the possibility of solving the problem of the diffraction of a plane wave by an oblate spheroid. If the spheroid is of non-vanishing thickness the boundary conditions lead to relations between the various components of the incident and scattered field which are quite intractable in view of the lack of orthogonality of vector wave functions and the appearance of certain non-separable factors which prevent effective use of the orthogonality of the scalar wave functions.

If the spheroid is taken of vanishing thickness the boundary conditions

are much simplified. In this case, although the vector solutions are again nonorthogonal, one can find a formal solution of the diffraction problem. The solution is a dubious one however in that it leads to edge singularities of too high an order and seems to lack uniqueness. These questions have been discussed by Boukamp and Meixner [2] but the exact nature of the difficulties does not appear to have been completely settled. This is rather surprising in view of the fact that singularities of the proper order and of the same order expected in the electromagnetic problem appear quite naturally in the corresponding scalar diffraction problem.

5. *Simple Boundary Problems*

We now turn to two problems which are closely related—the free oscillations of a spheroidal conductor and the oscillations of a spheroidal cavity. The problems differ only in that in the first case one uses the exterior or third kind radial functions while in the second case one must use the interior or first kind radial functions. The resonant frequencies of the cavity problem are of course real while those of the conductor problem are necessarily complex with the imaginary part representing the radiation damping.

To be definite let us again fix our attention on the oblate case. Consider the solution $\nabla \times \mathbf{r}\psi_{lm}$ where ψ_{lm} is a solution of the scalar wave equation in oblate spheroidal coordinates. One has

$$
\begin{aligned}
\nabla \times \mathbf{r}\psi_{lm} = \Bigg[& \boldsymbol{\eta}_1\left(\frac{-\xi}{[(\xi^2+\eta^2)(1-\eta^2)]^{1/2}}\right)\frac{\partial}{\partial\varphi} + \boldsymbol{\xi}_1\left(\frac{-\eta}{[(\xi^2+\eta^2)(1+\xi^2)]^{1/2}}\right)\frac{\partial}{\partial\varphi} \\
& + \boldsymbol{\phi}_1\left(\frac{[(1-\eta^2)(1+\xi^2)]^{1/2}}{\xi^2+\eta^2}\right)\left(\eta\frac{\partial}{\partial\xi} + \xi\frac{\partial}{\partial\eta}\right)\Bigg] u_{lm}(\eta)v_{lm}(\xi)\left.\begin{matrix}\sin\\ \cos\end{matrix}\right\} m\varphi
\end{aligned}
\tag{21}
$$

where the unit vectors η_1, ξ_1 and ϕ_1 have the directions indicated in the figure below.

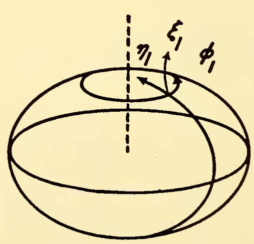


FIG. 1. The unit vectors for the oblate spheroid.

Let us consider the problem of trying to satisfy the boundary conditions of the cavity resonator or free oscillations on the spheroidal surface $\xi = \xi_0$.

If we assume that $\nabla \times \mathbf{r}\psi_{lm}$ represents the electric field then the boundary condition requires the tangential η and φ components to vanish. Setting $v_{lm}(\xi_0) = 0$ satisfies this condition for the η component. But to fulfill the condition on the φ component we must also set $d/d\xi\, v_{lm}(\xi_0) = 0$ and this condition cannot be met by the radial functions. If on the other hand we consider $\nabla \times \mathbf{r}\psi_{lm}$ to represent the magnetic field we again find that both $v_{lm}(\xi)$ and its normal derivative must vanish. Similar statements may be made about $\nabla \times \nabla \times \mathbf{r}\psi_{lm}$.

Part of the difficulty arises from the fact that both $\nabla \times \mathbf{r}\psi_{lm}$ and its curl contain all components of the field. In the spherical and cylindrical cases Maxwell's equations yield two distinct types of solutions—the TE and TM modes in which either the electric or magnetic component in the direction of propagation is absent. This happens in the spheroidal cases only in two special instances. The first is that in which variable ξ is zero and the second is the one in which the field is independent of the azimuthal variable φ. In the case of the limiting spheroid, $\xi = 0$, because the radius vector is tangent to the spheroid and in the η or negative η directions, $\nabla \times \mathbf{r}\psi_{lm}$ contains only the ξ and φ components, while $\nabla \times \nabla \times \mathbf{r}\psi_{lm}$ presents the full complement of components.

The cavity resonator problem obviously has no meaning for a spheroid of zero thickness but we may still discuss the free oscillation of a spheroidal conductor of zero thickness. By setting either $v_{lm}(0)$ or $(d/d\xi)\, v_{lm}(0)$ to zero we can make $\nabla \times \mathbf{r}\psi_{lm}$ satisfy boundary conditions of either the magnetic or electric type. As yet we have not investigated such oscillations in detail. Such an investigation should enable one to predict the resonance frequencies of the scattering cross section in the diffraction problem. In connection with the thin disk problem we may also consider solutions of the form $\nabla \times \mathbf{k}\psi_{lm}$. One has

$$
\begin{aligned}
\nabla \times \mathbf{k}\psi_{lm} = \Bigg[\mathbf{n}_1\left(\frac{-\eta}{[(\xi^2 + \eta^2)(1 - \eta^2)]^{1/2}}\right)\frac{\partial}{\partial\varphi} + \boldsymbol{\varsigma}_1\left(\frac{\xi}{[(\xi^2 + \eta^2)(1 + \xi^2)]^{1/2}}\right)\frac{\partial}{\partial\varphi} \\
+ \boldsymbol{\phi}_1\left(\frac{[(1 + \xi^2)(1 - \eta^2)]^{1/2}}{\xi^2 + \eta^2}\right)\left(\eta\frac{\partial}{\partial\eta} - \xi\frac{\partial}{\partial\xi}\right)\Bigg] u_{lm}(\eta)v_{lm}(\xi) \left.\begin{matrix}\sin\\ \cos\end{matrix}\right\} m\varphi.
\end{aligned}
\tag{22}
$$

If we try to use these fields to satisfy boundary conditions on a spheroid of finite size we find the same difficulty as previously mentioned, i.e.: both $v_{lm}(\xi_0)$ and $d/d\xi\, v_{lm}(\xi_0)$ are required to be zero. If, however, the field has azimuthal symmetry or if we deal with limiting spheroids we may again satisfy boundary conditions of the electric or magnetic type.

Solutions having azimuthal symmetry are formed by adding two non-solenoidal solutions of the vector wave equation:

$$
\mathbf{V}_1 = -\mathbf{i}\sin\varphi\, u_{l1}(\eta)v_{l1}(\xi) + \mathbf{j}\cos\varphi\, u_{l1}(\eta)v_{l1}(\xi) = \boldsymbol{\phi}_1 u_{l1}(\eta)v_{l1}(\xi). \tag{23}
$$

The result is clearly solenoidal. A second solution is obtained by taking the curl of this solution. One finds in the oblate system

$$\begin{aligned} \mathbf{V}_2 = \mathbf{n}_1\left(\frac{u_{l1}(\eta)}{(\xi^2+\eta^2)^{1/2}}\frac{\partial}{\partial\xi}\left[(1+\xi^2)^{1/2}v_{l1}(\xi)\right]\right) \\ - \xi_1\left(\frac{v_{l1}(\xi)}{(\xi^2+\eta^2)^{1/2}}\frac{\partial}{\partial\eta}\left[(1-\eta^2)^{1/2}u_{l1}(\eta)\right]\right). \end{aligned} \tag{24}$$

These wave functions have rather useful properties. We note that not all the components appear in both solutions which simplifies the boundary value problem. Although they are not individually orthogonal in respect to integration over a spheroidal surface, one can make good use of the orthogonality of the scalar wave functions. The magnetic and electric boundary conditions of the free oscillations of the spheroid or of a spheroidal cavity resonator can be simply satisfied by setting either $v_{l1}(\xi_0)$ or its derivative to zero.

The analog of the above fields in the prolate system has been the starting point of many investigations of the antenna problems.[3] Actually they suffice only to discuss the free oscillations of the antenna or a symmetrically excited transmitting antenna. In the case of a receiving antenna of non-vanishing thickness with its axis parallel to the electric vector the excitation is only approximately symmetric about the axis but the largest terms in the scattered field expansion are certainly the symmetric ones and one makes no great error by neglecting fields of higher symmetry when imposing the boundary conditions.

In connection with the oblate case one has a quite different group of problems which can profitably employ the symmetric field. One of these which has been solved recently [3] is that of an antenna on the axis of a circular disk which represents a finite ground plane. This problem has a rather special interest as the actual calculations were done for an oblate spheroid of vanishing thickness. The edge singularities which are currently blamed for the difficulties of the plane wave diffraction problem failed to give difficulty.

Other problems which appear quite feasible in the oblate case but which have not yet been worked out in any detail are the diffraction of a circularly symmetric wave by an aperture, the radiation of flanged transmission lines and waveguides operating in circularly symmetric modes, and the circularly symmetric modes of the hyperbolic horn.

At this point it seems clear that the vector wave functions of the spheroidal systems are really satisfactory only for the description of circularly symmetric fields. While it may well be that the oblate wave functions will prove of value in discussing the diffraction of plane waves by a thin disc this seems by no means assured. We have contented ourselves with discussing the more or less conventional types of solutions thus far since we have felt that an examination of these was of first importance.

---

[3]See for example Page, L., *The electrical oscillations of a prolate spheroid.* Paper II. *Prolate spheroidal wave functions,* Physical Review, Volume 65, 1944, p. 98.

6. *Existence of Tangential Solutions*

We now consider the problems of trying to find other solutions. One of the difficulties we have previously mentioned is that except in the circularly symmetric case neither of the solutions are tangential to the spheroids. The question arises as to whether it might be possible to find by methods quite different than those indicated here, solutions which have the tangential property. The answer to this question appears to be negative. If one assumes that it is possible to have a field in which the component normal to the spheroid vanishes and then writes down the equations which the remaining components must satisfy one is led to differential equations which have no solution.

Let $P = (0, P_2, P_3)$, be a solution whose component normal to the spheroid $u_1 =$ const., is zero. Then if we require that P be solenoidal, we must have

$$\nabla \cdot \mathbf{P} = 0 \quad \text{and} \quad \nabla \times \nabla \times \mathbf{P} - k^2\mathbf{P} = 0. \tag{25}$$

If these equations are written in component form it can be shown that P_2 must satisfy the differential equations[4]

$$\frac{(u_2^2 + 1)}{u_1}\frac{\partial P_2}{\partial u_1} - \frac{(1 - u_2^2)}{u_2}\frac{\partial P_2}{\partial u_2} + 2P_2 = 0 \tag{26}$$

$$\begin{aligned}(u_1^2 + 1)\frac{\partial^2 P_2}{\partial u_1^2} + (1 - u_2^2) + \frac{(u_1^2 + u_2^2)}{(u_1^2 + 1)(1 - u_2^2)}\frac{\partial^2 P_2}{\partial u_3^2} + 2u_1\frac{\partial P_2}{\partial u_1} - 4u_2\frac{\partial P_2}{\partial u_2} \\ + \left[k^2(u_1^2 + u_2^2) + \frac{2u_2^2(u_1^2 + u_2^2) - (u_1^2 + 1)}{(u_1^2 + u_2^2)(1 - u_2^2)}\right]P_2 = 0.\end{aligned} \tag{27}$$

From (26), we see that P_2 must have the form $P_2 = (1/x)f(x/y)g(u_3)$ where $x = u_1^2 + 1$, $y = 1 - u_2^2$, and f is an arbitrary function. By direct substitution it can be seen that (27) has no solution of this form except in the cases (a) $k = 0$, (*b*) x or $y =$ const. Hence, in general no tangential solutions exist for the spheroid.

Case (b) implies a solution exists which is tangential to a given constant spheroid.

The procedure outlined above should lead to a general criterion for those surfaces which possess tangential solutions. We do not have the complete answer to this as yet. One set of sufficient conditions for such surfaces is that the metric coefficients must satisfy $h_1 = h_1(u_1)$, $h_2/h_3 = g(u_2)$. These conditions are, of course, satisfied by spherical surfaces. However, we have not constructed solutions for other cases. Assuming that it might be possible to find such solutions it is still doubtful whether they would be orthogonal over the particular surface to which they are tangent.

---

[4]These calculations are based on the oblate spheroid. Similar results can be obtained for the prolate case.

There is one further possibility that should be mentioned. It consists of representing the field in rectangular components with each of the components an infinite series of scalar wave functions. Such a representation has the considerable advantage that it allows one to make excellent use of the orthogonality of the scalar wave functions. In actual practice it is rather difficult to force the solenoidal requirement on such a representation. Moglich [4] has attempted to solve the diffraction problem by this method but it appears that his solution has the wrong type of edge singularity.

Bibliography

1. Hansen, W. W., *A new type of expansion in radiation problems,* Physical Review, Volume 47, 1935, p. 139. Hansen, W. W., and Beckerley, J. G., *Radiation from an aerial over a plane earth of arbitrary characteristics,* Physics, Volume 7, 1936, p. 220. Hansen, W. W., *Transformations useful in certain antenna calculations,* Journal of Applied Physics, Volume 8, 1937, p. 282.
2. Bouwkamp, C. J., *A note on singularities occurring at sharp edges in electromagnetic diffraction theory,* Physica, Volume 12, 1946, p. 467. Meixner, J., and Andrejewski, W., *Strenge Theorie der Beugung ebener electromagnetischer Wellen an der vollkommen leitenden Kreisscheibe und an der kreisförmigen Öffnung im vollkommen leitenden ebenen Schirm,* Annalen der Physik, Volume 7, 1950, p. 157.
3. Leitner, A., and Spence, R. D., *Effect of a circular groundplane on antenna radiation,* Journal of Applied Physics, Volume 21, 1950, p. 1001.
4. Möglich, F., *Beugungserscheinungen an Körpern von ellipsoidischer Gestalt,* Annalen der Physik, Volume 83, 1927, p. 609.

The W.K.B. Approximation as the First Term of a Geometric–Optical Series

By H. BREMMER
Philips Research Laboratories, N. V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands

1. *Introduction*

Applications of the W.K.B. approximation usually refer to the propagation of waves through an inhomogeneous medium. With respect to such problems it is often possible to interpret this approximation as the first term of an infinite series, each term of which represents waves that are produced by a particular number of reflections inside the medium.[1] In this paper we shall investigate this series with reference to the simplest application of the W.K.B. approximation. This application concerns the ordinary differential equation:

$$(d^2y/dx^2) + k^2(x)y = 0. \tag{1}$$

In scalar optics this equation describes the behaviour of a plane wave propagated perpendicularly to the stratifications of a medium whose refractive index $\mu(x) = k(x)/k_0$ depends exclusively on the coordinate x. The quantity $k(x)$ can be interpreted as the local value of the wave number $2\pi/\lambda(x)$. Equations of the type (1) may also occur in vectorial problems; e.g., (1) is satisfied by the amplitude of a Hertzian vector in the case of a spherically symmetric medium with variable dielectric constant and a magnetic permeability equal to unity.[2]

2. *Derivation of the W.K.B. Approximation of (1) from a Discontinuous Model*

In what follows we consider an inhomogeneous space $x > 0$ (with variable wave number $k(x)$) that is adjacent to a homogeneous space $x < 0$ with a constant wave number k_0 . Provisionally we replace the inhomogeneous space

---

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

---

[1]See H. Bremmer, Handelingen, Natuur-en Geneeskundig Congres, Nijmegen 1939, p.; 88 Philips Research Reports 4, p. 189, 1949; Physica, Volume 15, p. 593, 1949. The one-dimensional problem has also been attacked by R. Landauer in his thesis.

[2]Compare H. Bremmer, *Terrestrial Radio Waves*, Elsevier Publishing Co., Houston-Amsterdam, 1949, p. 138.

$x > 0$ by a set of homogeneous layers $0 < x < x_1$, $x_1 < x < x_2$, $x_2 < x < x_3$, $\cdots$ with the successive constant wave numbers k_1, k_2, k_3, $\cdots$ (see Figure 1).

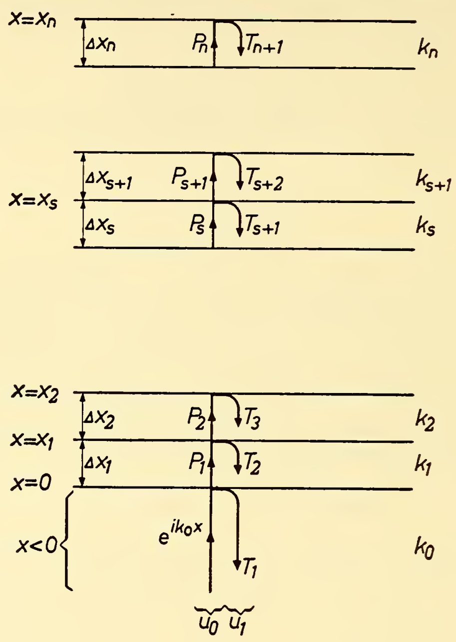


FIGURE 1

Later on we pass from this discontinuous medium to a continuously changing one by making the thicknesses $\Delta x_s = x_s - x_{s-1}$ of the various layers infinitely small.

We consider a plane wave $\exp \{i(k_0x - \omega t)\}$ arriving from the space $x < 0$ and travelling in the direction of increasing x; for convenience we term this direction "upwards," that of decreasing x "downwards." We shall investigate the behaviour of the wave $\exp \{ik_0x\}$ (the time factor $\exp \{-i\omega t\}$ will be omitted in the following discussions) when it enters into the discontinuous medium above the level $x > 0$. At the boundary $x = 0$ of the first layer the wave $\exp \{ik_0x\}$ is split into (*1*) a refracted wave P_1 penetrating into this layer and represented by $D_0 \exp \{ik_1x\}$, (*2*) a reflected wave T_1 returning to the space $x < 0$ and represented by $R_0 \exp \{-ik_0x\}$. The coefficients D_0 and R_0 are to be derived from the boundary conditions at $x = 0$. At any boundary we assume that u and du/dx shall be continuous. For the splitting of $\exp \{ik_0x\}$ into the waves P_1 and T_1 at $x = 0$ these conditions lead to the formulae:

$$R_0 = (k_0 - k_1)/(k_0 + k_1); \qquad D_0 = 2k_0/(k_0 + k_1).$$

The wave P_1 will be split at the next boundary $x = x_2$ into a refracted wave P_2 penetrating into the second layer and a refracted wave T_2 returning to the

first layer; these waves are proportional to $\exp\{ik_2x\}$ and $\exp\{-ik_1x\}$ respectively. This procedure of splitting is repeated at each next boundary. The boundary conditions for an arbitrary level $x = x_s$ lead to the following ratio of the amplitudes of P_s and P_{s+1} at this level:

$$(2) \qquad P_{s+1}(x_s)/P_s(x_s) = 2k_s/(k_s + k_{s+1}).$$

The chain of waves consisting of the sequence $P_1, P_2, P_3, \cdots$ may be termed the principal wave P. With the aid of (2) and the proportionality of this wave to $\exp\{ik_sx\}$ in the s-th layer we easily derive for the value of u_0 just below the level $x = x_n$:

$$\frac{2k_0}{k_0 + k_1}\exp\{ik_1\Delta x_1\}\frac{2k_1}{k_1 + k_2}\exp\{ik_2\Delta x_2\}\cdots\frac{2k_{n-1}}{k_{n-1} + k_n}\exp\{ik_n\Delta x_n\}.$$

With a view to the transition to a continuous medium, to be performed next, we write the latter expression in the alternative form:

$$(3) \qquad u_0(x_n - 0) = \exp\left\{-\sum_{s=0}^{n-1}\log(1 + \Delta k_s/2k_s) + i\sum_{s=1}^{n}k_s\Delta x_s\right\},$$

in which we have introduced the finite differences $\Delta k_s = k_{s+1} - k_s$ for the wave number. Passing to a continuous medium ($\Delta x_s \to 0$), the second sum in the exponent of (3) is transformed into the integral $i\int_{s=0}^{s=x_n} k(s)\,ds$, the first sum into

$$-\int_{s=0}^{s=x_n}\frac{dk_s}{2k_s} = -\frac{1}{2}\log\frac{k(x_n)}{k(0)}.$$

Thus we obtain the following formula for the principal wave in the continuous case:

$$u_0(x) = \exp\left\{-\frac{1}{2}\log(k(x)/k_0) + i\int_0^x k(s)\,ds\right\},$$

or

$$(4) \qquad u_0(x) = \left(\frac{k_0}{k(x)}\right)^{1/2}\exp\left\{i\int_0^x k(s)\,ds\right\}.$$

This expression just represents the W.K.B. approximation for the wave produced in the inhomogeneous space by the primary wave $\exp\{ik_0x\}$ arriving from the homogeneous space $x < 0$. Consequently we can give the following interpretation to this W.K.B. approximation: it represents the wave originating, by refractions, directly from the primary wave which arrives from the adjacent homogeneous space; its intensity is determined by the reflection processes which take place in any infinitely thin layer of the inhomogeneous medium.

3. *The First Correction Term to the W.K.B. Approximation*

The above derivation of the W.K.B. approximation suggests how to get additional contributions to the wave function. Returning to the discontinuous

model of Figure 1 we consider the reflected waves T_1, T_2, T_3, $\cdots$ that are generated by the principal P wave. These T-waves are travelling downwards, and cross the set of boundaries $x = x_s$ in a direction opposite to that of the P-wave. Each T-wave again produces a reflected wave (being a rising wave in this case) at any boundary. Owing to the inversion of the direction of propagation, the amplitude of a T-wave is multiplied by a factor $2k_{\sigma+1}/(k_{\sigma+1} + k_\sigma)$ when crossing the level $x = x_\sigma$ instead of the factor $2k_\sigma/(k_{\sigma+1} + k_\sigma)$ applying to the P-wave.

Let us consider the special wave T_{s+1} generated at the boundary $x = x_s$. This T-wave starts with an amplitude determined by the local value $u_0(x_s - 0)$ of the P-wave and by the reflection coefficient referring to a rising wave crossing the level x_s. Thus we obtain the following initial value of T_{s+1}:

$$T_{s+1}(x_s - 0) = u_0(x_s - 0)\,\frac{k_s - k_{s+1}}{k_s + k_{s+1}}. \tag{5}$$

The modifications undergone by T_{s+1} when travelling from $x = x_s$ to the lower level $x = x_m$ are of the same type as those encountered by the rising P-wave. Thus, by analogy to formula (3), we get the following ratio of the amplitudes of T_s at the levels mentioned:

$$\begin{aligned} &\frac{T_{s+1}(x_m + 0)}{T_{s+1}(x_s - 0)} \\ &\quad = \exp\left\{-\sum_{\sigma=m+2}^{s} \log\,(1 - \Delta k_{\sigma-1}/2k_\sigma) + i\sum_{\sigma=m+1}^{s} k_\sigma \Delta x_\sigma\right\}, \qquad x_m < x_s. \end{aligned} \tag{6}$$

The multiplication of (5) and (6) leads to a formula for $T_{s+1}(x_m + 0)$ in which the transition to a continuous medium may be very easily performed once more. The reflection coefficient occurring in (5), accordingly, is transformed into $-dk_s/2k_s$. In this way we arrive at the following amplitude at the level x for the wave that has been split off by reflection from the principal wave inside the infinitely thin layer $s < x < s + ds$:

$$-u_0(s)\,\frac{dk(s)}{2k(s)}\exp\left\{\frac{1}{2}\int_{\sigma=x}^{\sigma=s}\frac{dk(\sigma)}{k(\sigma)} + i\int_x^s k(\sigma)\,d\sigma\right\}$$

$$= -\frac{u_0(s)}{2[k(x)]^{1/2}}\,\frac{k'(s)}{[k(s)]^{1/2}}\,ds\,\exp\left\{i\int_x^s k(\sigma)\,d\sigma\right\}.$$

All the layers for which $s > x$, provide a similar contribution to the total wave u, which includes all the T-waves split off from the principal wave after one single reflection. The connection between this wave u_1 and the original P-wave u_0 thus proves to be given by

$$u_1(x) = -\frac{1}{2[k(x)]^{1/2}}\int_x^\infty ds\,\frac{k'(s)}{[k(s)]^{1/2}}\,u_0(s)\exp\left\{i\int_x^s k(\sigma)\,d\sigma\right\}. \tag{7}$$

4. *General Terms of the Complete Geometric-Optical Series; Recurrence Relations.*

As remarked in the preceding section each wave T_s generates a reflected wave (travelling upwards) when crossing any of the discontinuity levels of the model of Figure 1. These new reflected waves, U-waves, are thus produced after *two* successive reflections, one connected with the generation of the U-wave from T_s , the other one connected with the generation of the T_s-wave from the original P-wave. In their turn the U-waves produce further reflected waves which are thus produced after *three* various reflections. This reflection procedure is repeated ad infinitum. Therefore we get a complicated pattern of rays each of which is produced as the result of a definite number of reflections. Examples of such rays are shown in Figure 2 in which the number N attached to each ray refers to the number of reflections needed for the production of this very ray.

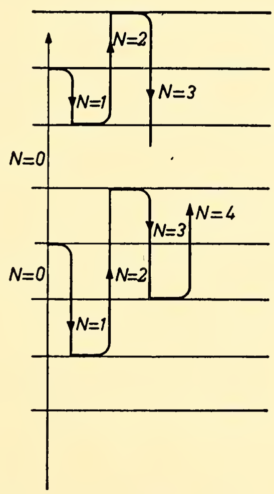


FIGURE 2

The N-classification here described still holds when passing to the limit of a continuously changing medium. For the latter we indicate by U_N the common contribution of all the waves produced by one and the same number of reflections N, the reflections in this case taking place in infinitely thin layers. With a view to this N-notation we have already marked the principal wave P by u_0 because this wave is produced without any reflection at all. Furthermore the system of the T-waves (which originate after one reflection) was accordingly marked u_1 in formula (7). Evidently all the terms u_{2N} with even subscripts $2N$ represent up-going waves, the terms u_{2N+1} with odd subscripts down-going waves.

The explicit expression for an arbitrary u_N is complicated while recurrence

formulae connecting two successive u_N-terms are rather simple. In this respect we recall that the derivation of u_1 in (7) is independent of the analytical form of the function $u_0(x)$. The same derivation therefore holds for the relation between any odd term u_{2N+1} and the preceding even term u_{2N}, the former being the result of the reflection losses of the latter. A similar relation can be derived for the dependence of the up-going u_{2N}-wave on the preceding down-going u_{2N-1}-wave. The two recurrence relations under consideration read explicitly:

$$u_{2N}(x) = \frac{1}{2[k(x)]^{1/2}} \int_0^x ds \frac{k'(s)}{[k(s)]^{1/2}} u_{2N-1}(s) \exp\left\{i \int_s^x k(\sigma)\, d\sigma\right\}, \tag{8a}$$

$$u_{2N+1}(x) = -\frac{1}{2[k(x)]^{1/2}} \int_x^\infty ds \frac{k'(s)}{[k(s)]^{1/2}} u_{2N}(s) \exp\left\{i \int_x^s k(\sigma)\, d\sigma\right\}. \tag{8b}$$

5. *The Complete Series $\Sigma\, u_N$ as a Solution of the Differential Equation*

In the discontinuous model the series $\sum_{N=0}^{\infty} u_N(x)$ represents the complete solution corresponding to the primary wave $\exp\{i\, k_0 x\}$ arriving from the homogeneous space $x < 0$. As a matter of fact we are sure not to have omitted anything at all when adding the contributions of all the rays that are possibly generated by any number of refractions and reflections at the discontinuity boundaries. This suggests that the same will hold in the limiting case of the continuous medium. We then have to verify that the differential equation (1) is satisfied by the series $y = \sum_{N=0}^{\infty} u_N$, the recurrence relations (8) being given.

This verification is performed as follows: We start by deriving new relations from a differentiation of (8) with respect to x, viz.,

$$u'_{2N} = -\left(\frac{k'}{2k} - ik\right)u_{2N} + \frac{k'}{2k} u_{2N-1}, \tag{9a}$$

$$u'_{2N+1} = -\left(\frac{k'}{2k} + ik\right)u_{2N+1} + \frac{k'}{2k} u_{2N}. \tag{9b}$$

Another differentiation of any of these identities, combined with an application of the other identity, leads to the following additional relation containing *three* consecutive terms

$$\frac{d^2u_N}{dx^2} + k^2u_N = \left(\frac{3}{4}\frac{k'^2}{k^2} - \frac{k''}{2k}\right)u_N + \left(\frac{k''}{2k} - \frac{k'^2}{k^2}\right)u_{N-1} + \frac{k'^2}{4k^2} u_{N-2}. \tag{10}$$

It is remarkable that here the discrimination between even and odd values of N has disappeared. The next step concerns the summation of (10) over $N = 2, 3, 4, \cdots$ while substituting

$$\sum_{N=0}^{\infty} u_N = y; \qquad \sum_{N=1}^{\infty} u_N = y - u_0; \qquad \sum_{N=2}^{\infty} u_N = y - u_0 - u_1.$$

The terms with u_1 can be expressed in u_0 with the aid of (9b) for $N = 0$ and of the differential quotient of this relation. The result of the summation mentioned then proves to be

$$\frac{d^2y}{dx^2} + k^2y = \frac{d^2u_0}{dx^2} + \frac{k'}{2k}\frac{du_0}{dx} + \left(k^2 - \frac{i}{2}k' - \frac{k'^2}{2k^2} + \frac{k''}{2k}\right)u_0 .$$

The right-hand member appears to be zero when evaluated according to (4). The series $y = \sum_{N=0}^{\infty} u_N$ (if the series for y'' is convergent) has thus been verified as a solution of the differential equation (1) in so far as the summation of (10) over N be legitimate (an example is given in section 8). The W.K.B. approximation now appears as the first term of an infinite series each term of which can be interpreted geometric-optically by the number of reflections needed for the production of the contributions represented by that term.

6. *Relations for the Total Rising and Downgoing Wave*

As remarked before, the terms u_N with even subscripts correspond to rising waves, those with odd subscripts to downgoing waves. Accordingly, we can split the complete solution $y = u$ into two parts, the *total* upgoing wave

(11a) $$u_\uparrow = u_0 + \epsilon^2 u_2 + \epsilon^4 u_4 + \cdots ,$$

and the *total* downgoing wave

(11b) $$u_\downarrow = \epsilon u_1 + \epsilon^3 u_3 + \epsilon^5 u_5 + \cdots ,$$

in which the parameter ϵ has to be taken equal to unity. This parameter is here introduced as a convenient expedient for the derivation of the several u_N-terms, as will be clear from the next sections. Evidently the power of ϵ, like the subscript N, indicates the number of reflections involved in the term under consideration.

The total up-going wave $u_\uparrow$ and the total downgoing wave $u_\downarrow$ satisfy a set of integral equations which lead to a rather simple survey of the consecutive u_N-terms. These integral equations are obtained at once from (8a) by multiplication by ϵ^{2N} and a summation over $N = 1, 2, 3, \cdots$ and from (8b) by multiplication by ϵ^{2N+1} and a summation over $N = 0, 1, 2, \cdots$. The equations in question are then found to be

(12a) $$u_\uparrow(x) - \frac{\epsilon}{2[k(x)]^{1/2}} \int_0^x ds \frac{k'(s)}{[k(s)]^{1/2}} u_\downarrow(s) \exp\left\{i \int_s^x k(\sigma)\, d\sigma\right\} = u_0(x),$$

(12b) $$u_\downarrow(x) + \frac{\epsilon}{2[k(x)]^{1/2}} \int_x^\infty ds \frac{k'(s)}{[k(s)]^{1/2}} u_\uparrow(s) \exp\left\{i \int_x^s k(\sigma)\, d\sigma\right\} = 0.$$

These integral equations are completely equivalent to the relations (8)

because the latter can be obtained from (12) by equating to zero the terms occurring with one and the same power of ϵ after substituting (11).

It is also possible to eliminate either of the functions $u_\uparrow(x)$ or $u_\downarrow(x)$ from (12) in order to obtain a *single* integral equation of the Volterra type for the other function. The development of $u_\uparrow$ or $u_\downarrow$ into powers of ϵ proves to be identical with the Neumann-Liouville expansion of the solution of any of these single integral equations with the aid of iterated kernels. Thus our geometric-optical splitting can be reduced mathematically to a well-known method for solving integral equations.

We conclude this section with the derivation of some other relations concerning the functions $u_\uparrow$ and $u_\downarrow$. For this purpose we multiply (9a) by ϵ^{2N} and make a summation over $N = 1, 2, 3 \cdots$ and also multiply (9b) by ϵ^{2N+1} while summing over $N = 0, 1, 2 \cdots$. An addition of the resulting identities, using moreover the relation $u = u_\uparrow + u_\downarrow$, leads to the relations in question:

$$u_\uparrow = \frac{u}{2} + \frac{u' + (1-\epsilon)\dfrac{k'}{2k}u}{2ik},$$

(13)

$$u_\downarrow = \frac{u}{2} - \frac{u' + (1-\epsilon)\dfrac{k'}{2k}u}{2ik}.$$

We recall the introduction of ϵ as an expedient for the derivation of relations between the u_N terms. The actual value $\epsilon = 1$ reduces (13) to the simple formulae

$$u_\uparrow = \frac{u}{2} + \frac{u'}{2ik},$$

(14)

$$u_\downarrow = \frac{u}{2} - \frac{u'}{2ik}.$$

The influence of the inhomogeneity of the medium on the applicability of the W.K.B. approximation is shown very clearly by finally deriving the following identities from a differentiation of (14) and an application of (1):

$$\frac{d}{dx}\left[k^{1/2}\exp\left\{-i\int_0^x k(s)\,ds\right\}u_\uparrow(x)\right] = \frac{k'}{2k} \times k^{1/2}\exp\left\{-i\int_0^x k(s)\,ds\right\}u_\downarrow(x),$$

(15)

$$\frac{d}{dx}\left[k^{1/2}\exp\left\{i\int_0^x k(s)\,ds\right\}u_\downarrow(x)\right] = \frac{k'}{2k} \times k^{1/2}\exp\left\{i\int_0^x k(s)\,ds\right\}u_\uparrow(x).$$

In fact, a small inhomogeneity indicates small values of k' and a possible neglection of the right-hand members which leads to the W.K.B. approximation for $u_\uparrow$ and $u_\downarrow$. The deviation from the W.K.B. values is directly dependent on the numerical values of the right-hand members of (15).

7. *Transition to Other Variables instead of u and x*

The determination of the u_N-terms is considerably facilitated by the introduction of the quantities

$$\xi = \int_0^x k(s)\, ds \qquad \text{and} \qquad I(x) = [k(x)]^{1/2}u(x)$$

instead of x and u respectively. The terms of the series $I(x) = \sum_{N=0}^{\infty} \epsilon^N I_N(x)$, which corresponds to the original series $u(x) = \sum_{N=0}^{\infty} \epsilon^N u_N(x)$ are then simply obtained as follows: the solution of the differential equation

$$\frac{d^2I}{d\xi^2} + \left\{1 - \epsilon \frac{dR\,(\xi)}{d\xi} - \epsilon^2 R^2(\xi)\right\} I = 0 \tag{16}$$

which satisfies the boundary conditions

$$\begin{aligned} &I(1 + i\epsilon R) - i\frac{dI}{d\xi} = 2(k_0)^{1/2} && \text{at } \xi = 0, \\ &I(1 - i\epsilon R) + i\frac{dI}{d\xi} = 0 && \text{at infinity,} \end{aligned} \tag{17}$$

is developed with respect to ϵ (the coefficient of ϵ^N yielding I_N); the new parameter R is defined by

$$R = \frac{dk/dx}{2k^2} = \frac{dk/d\xi}{2k}.$$

The correctness of this procedure is demonstrated as follows: The differential equation (16) is derived by multiplying (10) by ϵ^N, by summing over $N = 2, 3, \cdots$ and final transition from the variables u and x to I and ξ. Further, in virtue of (13) the transposition of the boundary conditions (17) to the original variables u and x, simply reads

$$u_\uparrow = u_0 = 1 \qquad \text{at} \qquad x = 0, \tag{18a}$$

$$u_\downarrow = 0 \qquad \text{at} \qquad x = \infty. \tag{18b}$$

The first condition (18a) states the vanishing at $x = 0$ of the difference of the total rising wave and the principal wave u_0. The contributions to this difference, as observed at a special level x, are generated in the space between the x level under consideration and the boundary at $x = 0$; this difference therefore vanishes if the space mentioned is reduced to a zero thickness, i.e. if the x level approaches the boundary $x = 0$. The other condition (18b) is verified by the corresponding property at infinity, namely the vanishing there of the total downgoing wave.

It is to be noted that the variable ξ is proportional to the number of wavelengths

$$\int_0^x \frac{ds}{\lambda(s)} = \frac{1}{2\pi} \int_0^x k(s)\, ds$$

comprised between $x = 0$ and the level under consideration. In other words, the introduction of ξ amounts to taking the local values of the wavelengths as units when measuring the distances x. Finally the quantity $R(x)$ can be interpreted as the reflection coefficient of a layer with thickness $1/k(x) = \lambda(x)/2\pi$.

8. *An Example of the Geometric-Optical Series*

The preceding theory may be illustrated by the inhomogeneous medium with a constant value $-R_0$ of $R(\xi)$. The equation (16) here reduces to

$$\frac{d^2 I}{d\xi^2} + (1 - \epsilon^2 R_0^2) I = 0.$$

The solution satisfying the boundary conditions (17) reads

$$I = i(k_0)^{1/2} \frac{1 - i\epsilon R_0 - (1 - \epsilon^2 R_0^2)^{1/2}}{\epsilon R_0} \exp \{i\xi(1 - \epsilon^2 R_0^2)^{1/2}\} \qquad \xi > 0, \tag{19}$$

if we assume ξ complex with a small positive argument so that $e^{i\xi} \to 0$ and $e^{-i\xi} \to \infty$ at infinity (slightly absorbing medium). The function $R(\xi) = -R_0$ corresponds to the situation:

$$k(x) = \begin{cases} k_0\,, & x < 0 \\ \dfrac{k_0}{1 + 2k_0 R_0 x}, & x > 0, \end{cases}$$

which is shown schematically in Figure 3.

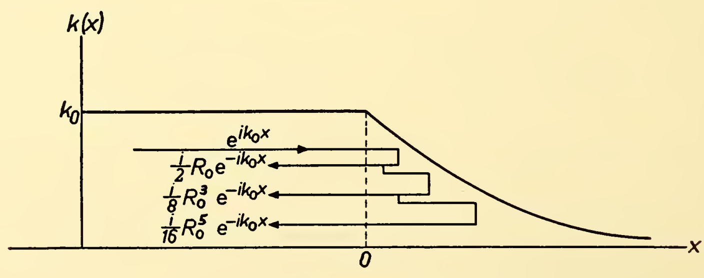


FIGURE 3

The primary wave $\exp \{ik_0 x\}$ arriving from the homogeneous space $x < 0$ produces reflected waves in the inhomogeneous space $x > 0$. These reflected

waves reach the homogeneous space in which we can represent their sum by a function of the form $u(x) = T \exp\{-ik_0x\}$. The quantity T acts as the *resulting* reflection coefficient of the complete inhomogeneous space extending from $x = 0$ to $x = \infty$. The u-value of the total field at $x = 0-$, viz. $1 + T$, must be equal to the u-value at $x = 0+$, viz. $I(0)/(k_0)^{1/2}$. In virtue of (19) this boundary condition leads to the following value of T:

$$T = i\,\frac{1 - (1 - \epsilon^2R_0^2)^{1/2}}{eR_0}.$$

The expansion

$$T = -i \sum_{N=1}^{\infty} (-1)^N \binom{1/2}{N} (\epsilon R_0)^{2N-1}$$

shows the distribution of the amplitudes of the total waves that are generated by 1, 3, 5, $\cdots$ reflections in the inhomogeneous space. Indeed, according to the above theory these amplitudes are equal to the coefficients of ϵ^1, ϵ^3, ϵ^5 $\cdots$. The first few reflected waves are thus given by the following expressions

1 reflection: $$i\binom{1/2}{1}R_0 \exp\{-ik_0x\} = \frac{i}{2}\,R_0 \exp\{-ik_0x\},$$

2 reflections: $$-i\binom{1/2}{2}R_0^3 \exp\{-ik_0x\} = \frac{i}{8}\,R_0^3 \exp\{-ik_0x\},$$

3 reflections: $$i\binom{1/2}{3}R_0^5 \exp\{-ik_0x\} = \frac{i}{16}\,R_0^5 \exp\{-ik_0x\}.$$

The amplitudes of the waves reflected to the homogeneous space are thus the binomial coefficients of $\frac{1}{2}$ in this example. At the same time we infer the convergence of the splitting procedure if $\epsilon R_0 = R_0 < 1$.

As for the convergence of the u_N-series in other examples, the well-known insufficiency of the W.K.B. approximation in the neighbourhood of zeros of $k(x)$ can be interpreted in our theory as a very slow convergence of the u_N-series.

Remarks Concerning Wave Propagation in Stratified Media

By S. A. SCHELKUNOFF
Bell Telephone Laboratories

1. *Wave Equations*

Under a great variety of conditions the problem of wave propagation reduces to one or more equations of the form

$$\frac{d^2u}{dx^2} + F(x)u = 0. \tag{1}$$

The complete wave function is $u \exp \{i\omega t\}$, where t is the time and ω is the frequency in radians per second. If $F(x)$ is not analytic, u and its first derivative must be continuous. In the case of uniform plane waves of frequency ω, incident on a horizontally stratified layer, (Figure 1) u is either the electric intensity E

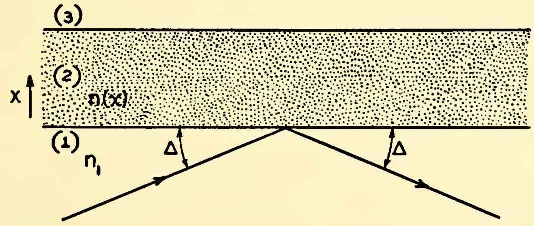


FIG. 1. A uniform plane wave incident on an inhomogeneous layer.

or the magnetic intensity H according as E or H is parallel to the layer. If the angle of elevation Δ is small,

$$F(x) = \omega^2\mu\epsilon_1[2(n - n_1) + \Delta^2], \tag{2}$$

where μ, ϵ, n are respectively the permeability, the dielectric constant and the index of refraction. The subscripts refer to the homogeneous medium below the stratified medium. It is assumed that μ is constant throughout the entire medium.

Equation (1) is not the primary wave equation in the sense that, in general, it does not arise directly from the physical laws and does not express the physical

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

conditions completely. For example, if E is parallel to the layer and H_t is the component of H parallel to the layer, we obtain

$$\text{(3)} \qquad \frac{dE}{dx} = -i\omega\mu\, H_t\,, \qquad \frac{dH_t}{dx} = -i\omega\epsilon_1[2(n - n_1) + \Delta^2]E.$$

Since μ is independent of x, we may reduce (3) to (1) by eliminating H_t. In the more general case we have a system of two first order wave equations with two wave functions

$$\text{(4)} \qquad \frac{du}{dx} = -if(x)v, \qquad \frac{dv}{dx} = -ig(x)u.$$

If either $f(x)$ or $g(x)$ is not analytic, u and v are required to be continuous. If either $f(x)$ or $g(x)$ is independent of x, then either u or v satisfies (1) and the corresponding boundary conditions. But even in this case there are some advantages in dealing directly with (4) rather than with (1). In any case, the proper definition of the reflection coefficient requires both wave functions u, v, even though one of them may be the derivative of the other.

2. *Definition of the Reflection Coefficient*

In a homogeneous medium $f(x)$ and $g(x)$ are constants and the general solution may be expressed as the sum of two progressive wave functions

$$\text{(5)} \qquad u = Ae^{-i\beta_1 x} + Be^{i\beta_1 x}, \qquad v = K_1^{-1}(Ae^{-i\beta_1 x} - Be^{i\beta_1 x}),$$

where

$$\text{(6)} \qquad \beta_1 = (f_1 g_1)^{1/2}, \qquad K_1 = (f_1/g_1)^{1/2}.$$

The parameter K_1 equals the ratio u/v for the wave traveling in the positive x direction and $-u/v$ for the wave traveling in the opposite direction.

If the source of the wave is at $x = -\infty$ and if the medium is discontinuous at $x = 0$, the ratio $q = B/A$ is called the coefficient of reflection for the u-function. Expressing q in terms of $u(0)$ and $v(0)$, we have

$$\text{(7)} \qquad q = \frac{Z - K_1}{Z + K_1}, \qquad Z = \frac{u(0)}{v(0)}.$$

Thus, the reflection coefficient depends solely on the ratio of the wave functions at the discontinuity and on the parameter K_1 of the homogeneous medium containing the incident wave.

In general, no meaning can be attached to the reflection coefficient when the incident wave is in an inhomogeneous medium for the simple reason that in general we cannot decompose the total wave into "progressive" components. A *formal* expression such as

$$\text{(8)} \qquad u = A(x)\, e^{-i\Phi(x)},$$

in the form of an apparently "progressive wave" function in a homogeneous medium does not insure that it represents a wave with "progressive" physical characteristics. Consider, for instance, the following wave function in a homogeneous medium

$$u = \cos \beta x + 0.01\, e^{-i\beta x}. \tag{9}$$

This function may be expressed in the form (8); thus

$$A(x) = [(1.01)^2 - 1.02 \sin^2 \beta x]^{1/2}, \qquad \Phi(x) = \tan^{-1}\left(\frac{1}{101} \tan \beta x\right). \tag{10}$$

The phase function $\Phi(x)$ is a monotonic increasing function. And yet (9) represents primarily a standing wave.

Thus, in general, we shall restrict the definition of the reflection coefficient given by (7) to conditions such as those expressed diagrammatically in Figure 2(a), where the incident wave travels in a homogeneous medium (1) and im-

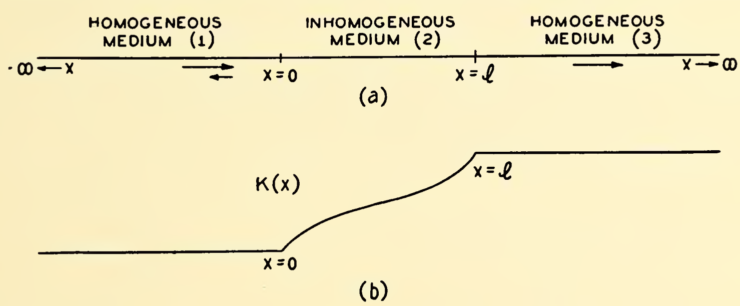


Fig. 2. A diagram of reflection from an inhomogeneous layer (a) and a function governing transmission characteristics of the entire medium (b).

pinges on an inhomogeneous layer (2) of thickness l. From $x = l$ to $x = \infty$ we have assumed another homogeneous medium to enable us to recognize the transmitted wave. Under some conditions we may be able to let l approach infinity. As we see from equation (7) the reflection phenomena depend on the properties of the function

$$K(x) = [f(x)/g(x)]^{1/2} \tag{11}$$

which for equation (1) becomes

$$K(x) = [F(x)]^{1/2} \qquad \text{or} \qquad K(x) = [F(x)]^{-1/2}, \tag{12}$$

according as $g(x)$ or $f(x)$ is independent of x.

3. *Reflection from Infinite Layers*

It is not easy to find a simple exact expression for the reflection coefficient from a finite inhomogeneous layer. There is no reason, of course, why we should

not be content with good approximations. But there appears to be something in human nature, or more probably in the intellectual habits acquired in childhood, that makes one yearn for the "exact" answer to a given problem. As far as this writer is concerned, the exact answer is merely a specification of fairly easy successive steps by which one can obtain increasingly better approximations until further improvement loses its meaning. We except, of course, those trivial cases in which the exact answers are given by integers. And in practice it makes little difference whether a given problem is solved approximately or replaced by an approximating problem which is then solved exactly.

Nevertheless in the case of wave propagation there is a class of idealized problems which at first sight appear to resemble closely important practical problems and which can be solved in closed form in terms of elementary functions. There is a temptation to apply these solutions to the practical problems, and we wish to warn the readers against such applications. For instance, if the dielectric constant is given by[1]

$$\epsilon(x) = \epsilon_1 + e^{\xi}(e^{\xi} + 1)^{-2}[(\epsilon_2 - \epsilon_1)(e^{\xi} + 1) + \epsilon_3], \qquad \xi = 2\pi x/h, \tag{13}$$

equations (3) may be solved in terms of hypergeometric functions and a simple formula may be obtained for the reflection coefficient at $x = -\infty$. If x varies from $-\infty$ to ∞, $\epsilon(x)$ varies from ϵ_1 to ϵ_2. If $-|\epsilon_2 - \epsilon_1| \leq \epsilon_3 \leq |\epsilon_2 - \epsilon_1|$, the

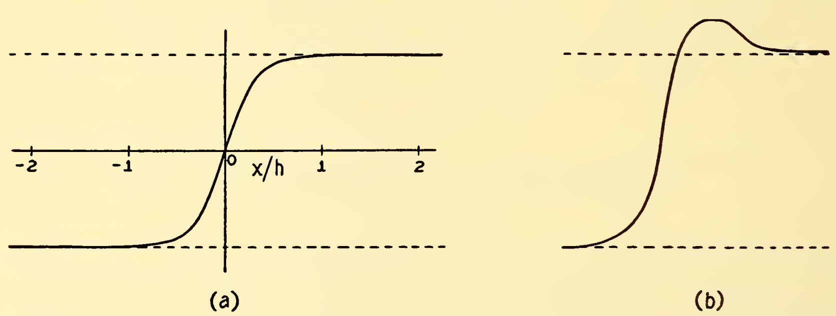


FIG. 3. Types of functional characteristics of an inhomogeneous layer for which exact solutions of the wave equation are known.

variation is monotonic; otherwise, ϵ first rises or falls to a maximum or a minimum and then goes back to the final value (Figure 3b). If $\epsilon_3 = 0$,

$$\epsilon(x) = \tfrac{1}{2}(\epsilon_2 + \epsilon_1) + \tfrac{1}{2}(\epsilon_2 - \epsilon_1) \tanh \tfrac{1}{2}\xi \tag{14}$$

and the transition is antisymmetric about $x = 0$ as shown in Figure 3a. If $\epsilon_2 = \epsilon_1$,

$$\epsilon(x) = \epsilon_1 + \tfrac{1}{4}\,\epsilon_3 \operatorname{sech}^2 \tfrac{1}{2}\xi, \tag{15}$$

[1] P. S. Epstein, *Reflection of waves in an inhomogeneous absorbing medium,* Proceedings of the National Academy of Sciences, Volume 16, 1930, pp. 627–637.

and we have a symmetric ridge or a valley in the distribution of $\epsilon(x)$. The parameter ξ expresses roughly the thickness of the layer, but actually the dielectric constant attains its boundary values ϵ_1, ϵ_2 only at $x = -\infty$, $x = +\infty$ respectively.

In spite of the apparent resemblance between the Epstein layer (Figure 3a) and a finite layer of thickness h, the reflection coefficients in the two cases are of different orders of magnitude if the thickness of the atmospheric layer is large compared with the vertical wavelength,

$$\lambda_x = \lambda/\Delta, \tag{16}$$

where Δ is the angle between the incident rays and the layer, as indicated on Figure 1. It is known[2] that the reflection coefficient for the Epstein layer varies as $\exp\{-2\pi h/\lambda_x\}$; we shall show that for a finite layer it varies only as $(2\pi h/\lambda_x)^{-n}$ where for small Δ the exponent n equals unity or two (see equations 37 and 39). It is only at normal incidence that n may be fairly large; but it is still true that the orders of magnitude of the reflection coefficients for thick layers are quite different in the two cases. The explanation is that as $\lambda_x/h \to 0$ the infinite regions from $(-\infty, h)$ and (h, ∞) have an increasingly pronounced effect on the reflection coefficient at $x = -\infty$. In the language of the electrical engineer the various sections of the medium become increasingly better "matched" to each other and reflections are greatly reduced.

If $h \ll \lambda_x$ the two coefficients are substantially the same, for as h/λ_x approaches zero the reflection coefficient approaches a value depending on the relative difference between the initial and final values of K and not on the manner of transition.

4. *Reduction of the Wave Equations to a Canonical Form*

Let us transform (4) by introducing a new dimensionless independent variable

$$\vartheta(x) = \int_a^x [f(x)g(x)]^{1/2}\, dx \tag{17}$$

and new wave functions $U(x)$, $V(x)$

$$U(x) = [K(x)]^{-1/2}u(x), \qquad V(x) = [K(x)]^{1/2}v(x), \tag{18}$$

where $K(x)$ is defined by (11). Thus we obtain

$$\frac{dU}{d\vartheta} = -iV - \frac{K'}{2K}\,U, \qquad \frac{dV}{d\vartheta} = -iU + \frac{K'}{2K}\,V, \tag{19}$$

where the prime denotes differentiation with respect to the phase variable ϑ.

[2] *Ibid.*

5. *The Case of Zero Reflection*

If

$$K = \text{const.}, \tag{20}$$

the solutions of (19) and (4) are respectively

$$\begin{aligned} U &= Ae^{-i\vartheta} + Be^{i\vartheta}, \qquad & V &= Ae^{-i\vartheta} - Be^{i\vartheta}, \\ u &= K^{1/2}(Ae^{-i\vartheta} + Be^{i\vartheta}), \qquad & v &= K^{-1/2}(Ae^{-i\vartheta} - Be^{i\vartheta}). \end{aligned} \tag{21}$$

Irrespective of the dependence of the phase variable ϑ on x, the ratio of the forward progressive wave functions is K and we have no reflections. This is the only case in which there are no reflections. In the case of a stratified atmosphere either $f(x)$ or $g(x)$ is independent of x; hence $K(x)$ depends on x and we must have reflections.

It has been suggested that there is a case of variable F in (1) and hence of variable K in which there are no reflections. We note that (1) is a special case of (4) in which

$$v = i\frac{du}{dx}, \qquad f(x) = 1, \qquad g(x) = F(x), \qquad K(x) = [F(x)]^{-1/2}. \tag{22}$$

Eliminating V from (19), we find

$$\frac{d^2U}{d\vartheta^2} = -U + \left[\frac{3(K')^2}{4K^2} - \frac{K''}{2K}\right]U. \tag{23}$$

The bracketed expression vanishes if

$$K(\vartheta) = (P\vartheta + Q)^{-2}, \tag{24}$$

where P and Q are constants of integration. In this case,

$$U = Ae^{-i\vartheta} + Be^{i\vartheta}, \qquad u = \frac{Ae^{-i\vartheta} + Be^{i\vartheta}}{P\vartheta + Q}. \tag{25}$$

Thus two seemingly progressive waves appear to exist independently in the medium for which K is given by (24).

However, the reflection coefficient (7) depends on the ratio u/v. The second wave function in the present case is

$$\begin{aligned} V &= A\left(1 - \frac{iP}{P\vartheta + Q}\right)e^{-i\vartheta} - B\left(1 + \frac{iP}{P\vartheta + Q}\right)e^{i\vartheta}, \\ v &= A(P\vartheta + Q - iP)e^{-i\vartheta} - B(P\vartheta + Q + iP)e^{i\vartheta}. \end{aligned} \tag{26}$$

For a "progressive" wave moving in the positive ϑ direction we have

$$\frac{u(\vartheta)}{v(\vartheta)} = [(P\vartheta + Q)^2 - iP(P\vartheta + Q)]^{-1}. \tag{27}$$

Hence, if the medium is homogeneous in the interval $(-\infty, \vartheta_0)$ and then inhomogeneous, we have

$$q = -\frac{P}{P + 2i(P\vartheta_0 + Q)}\,. \tag{28}$$

The reflection coefficient vanishes only at $\vartheta_0 = -\infty$.

It is instructive to consider also the complex power flow. If u is the electric intensity and v the magnetic intensity of an electromagnetic wave, the power flow in the direction normal to the stratification is

$$W = \tfrac{1}{2}\, uv^*. \tag{29}$$

For the "progressive" wave in the above case we have

$$W = \frac{1}{2}\left(1 + \frac{iP}{P\vartheta + Q}\right) AA^*. \tag{30}$$

In the vicinity of $\vartheta_0 = -Q/P$ the reactive power flow is very large. At $\vartheta = \vartheta_0$ the power flow is infinite and this point is an effective barrier to the incoming wave. As equation (28) indicates, the reflection is total.

6. *First Order Reflection Coefficient*

To the extent to which $K'/2K$ is negligible compared with unity, equations (21) represent approximate solutions of our wave equations. In order to obtain the first order reflection coefficient we convert (19) into the following set of integral equations

$$\begin{aligned} U(\vartheta) &= U_0(\vartheta) - \int_a^\vartheta \frac{K'(\varphi)}{2K(\varphi)}\,[U(\varphi)\cos(\vartheta - \varphi) + iV(\varphi)\sin(\vartheta - \varphi)]\,d\varphi, \\ V(\vartheta) &= V_0(\vartheta) + \int_a^\vartheta \frac{K'(\varphi)}{2K(\varphi)}\,[V(\varphi)\cos(\vartheta - \varphi) + iU(\varphi)\sin(\vartheta - \varphi)]\,d\varphi, \end{aligned} \tag{31}$$

where U_0 and V_0 are of the same form as U and V in (21). Let us assume that the inhomogeneous medium extends from $\vartheta = 0$ to $\vartheta = \Theta$ and that for $\vartheta > \Theta$ we have

$$U(\vartheta) = e^{-i\vartheta}, \qquad V(\vartheta) = e^{-i\vartheta}. \tag{32}$$

If $K'/2K \ll 1$, then in the first approximation these expressions hold also for $\vartheta < \Theta$. To obtain the next approximation we substitute (32) in the integrands of (31). Hence, letting $a = \Theta$, we have

$$\begin{aligned} U(\vartheta) &= e^{-i\vartheta} - e^{i\vartheta} \int_\Theta^\vartheta \frac{K'(\varphi)}{2K(\varphi)}\, e^{-2i\varphi}\, d\varphi \\ V(\vartheta) &= e^{-i\vartheta} + e^{i\vartheta} \int_\Theta^\vartheta \frac{K'(\varphi)}{2K(\varphi)}\, e^{-2i\varphi}\, d\varphi. \end{aligned} \tag{33}$$

For the U function the coefficient of reflection at $\vartheta = 0$ is

$$q = \int_0^\Theta \frac{K'(\varphi)}{2K(\varphi)} e^{-2i\varphi}\, d\varphi = \frac{1}{2}\int_0^\Theta \frac{d \log K(\varphi)}{d\varphi} e^{-2i\varphi}\, d\varphi. \tag{34}$$

It is to be noted that Θ is the thickness of the layer in radians. If $2\Theta \ll 1$, then

$$q \simeq \tfrac{1}{2} \log (K_2/K_1), \qquad K_1 = K(0), \qquad K_2 = K(\Theta). \tag{35}$$

If $\log K$ varies linearly with ϑ (Figure 4)

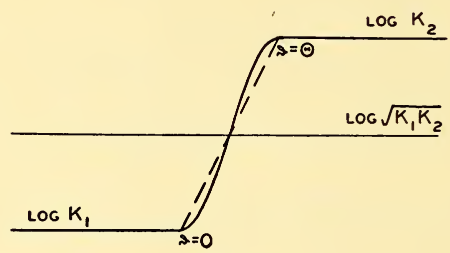


FIG. 4. Two simple transition characteristics of an inhomogeneous layer: (a) linear, (b) antisymmetric cubic with a continuous tangent.

$$\log K = \log K_1 + \frac{\vartheta}{\Theta} \log (K_2/K_1), \tag{36}$$

then

$$q = \frac{1}{2} \log (K_2/K_1) \frac{\sin \Theta}{\Theta} e^{-i\Theta}. \tag{37}$$

If $\log K$ is an antisymmetric cubic in $(\vartheta - \frac{1}{2}\Theta)$ (Figure 4), and if the first derivative is continuous at the boundaries of the layer,

$$\begin{aligned} \log K = \log (K_1K_2)^{1/2} &+ \frac{3}{2\Theta}\left(\vartheta - \frac{1}{2}\Theta\right) \log (K_2/K_1) \\ &- \frac{2}{\Theta^3}\left(\vartheta - \frac{1}{2}\Theta\right)^3 \log (K_2/K_1). \end{aligned} \tag{38}$$

In this case

$$q = \frac{3 \log (K_2/K_1)}{2\Theta^2}\left(\frac{\sin \Theta}{\Theta} - \cos \Theta\right) e^{-i\Theta}. \tag{39}$$

As Θ approaches zero, (37) and (39) approach (35). As Θ approaches infinity, the reflection coefficients in these cases vary respectively as $1/\Theta$ and $1/\Theta^2$. If K and its first n derivatives are continuous at the boundaries of the layer but the $(n + 1)$-th derivative is discontinuous, then for sufficiently large Θ the reflection coefficient varies as $1/\Theta^{n+1}$. This is true irrespective of the precise functional dependence of K on ϑ. A "rapid" change in the $(n + 1)$-th

derivative occurs when most of the change takes place in an interval small compared with the radian; such a change is equivalent to a discontinuity in this derivative. All this may be deduced from (34) if we integrate it by parts repeatedly.

If we substitute (33) in the integrands of (31), we obtain the third terms in the approximate series for U and V,

$$U_2(\vartheta) = e^{-i\vartheta} \int_\Theta^\vartheta \frac{K'(\varphi_1)}{2K(\varphi_1)} e^{2i\varphi_1}\, d\varphi_1 \int_\Theta^{\varphi_1} \frac{K'(\varphi)}{2K(\varphi)} e^{-2i\varphi}\, d\varphi, \tag{40}$$

$$V_2(\vartheta) = U_2(\vartheta).$$

Hence, the second approximation to the reflection coefficient is

$$q_1 = \frac{q}{1+\chi}, \tag{41}$$

where q is given by (34) and χ is the value of $U_2(\vartheta)$ for $\vartheta = 0$. The sequence of successive approximations can be continued and q can be expressed as a ratio of two series; but if the first order reflection coefficient is not small compared with unity, it is probably better to use the method described in the following section.

7. *Calculation of Strong Reflections*

Reflections increase as K'/K increases. Since $f(x)$ and $g(x)$ are proportional to frequency, ϑ is proportional to frequency; hence, K'/K is inversely proportional to frequency. For this reason the longer waves are reflected more strongly than the shorter waves. Reflections depend on the rate of change of K and on the total change in K across the layer. If the total change in K is small, the reflection coefficient is small even if its rate of change per radian is large. In the troposphere strong reflections occur near grazing incidence, where the wavelength (16) in the direction normal to the layer is relatively large even though the wavelength in the direction of propagation may be small. In such a case the tropospheric layers are relatively thin in the sense that the phase of the wave does not change much as we pass across the layer in the normal direction. When the electrical engineer is faced with a similar situation in the case of tapered transmission lines, he replaces them by a chain of lumped parameters. Since his special language may not be familiar to every reader of this paper, we shall translate his method into mathematical terms.

Suppose that the medium described by equations (4) extends from $x = 0$ to $x = l$. Let

$$s = l/n \tag{42}$$

be a fraction of the total interval so small that no matter where we select a subinterval of length s, the change in the phase variable (17) does not exceed,

let us say, $\pi/4$. That is, s is equal to or less than one eighth of the shortest wavelength in the x direction. Let us integrate the first equation of the set (4) in the following subintervals: $(0, \frac{1}{2}s)$, $(\frac{1}{2}s, \frac{3}{2}s)$, $(\frac{3}{2}s, \frac{5}{2}s)$, $\cdots$ $(l - \frac{1}{2}s, l)$. Similarly, let us integrate the second equation of the set in the intervals $(0, s)$, $(s, 2s)$, $(2s, 3s)$, $\cdots$ $(l - s, l)$. Thus we obtain

$$
\begin{aligned}
u(0) &= u(\tfrac{1}{2}s) + i\int_0^{\frac{1}{2}s} f(x)v(x)\,dx, & v(0) &= v(s) + i\int_0^{s} g(x)u(x)\,dx,\\
u(\tfrac{1}{2}s) &= u(\tfrac{3}{2}s) + i\int_{\frac{1}{2}s}^{\frac{3}{2}s} f(x)v(x)\,dx, & v(s) &= v(2s) + i\int_s^{2s} g(x)u(x)\,dx,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
u(l - \tfrac{1}{2}s) &= u(l) + i\int_{l-\frac{1}{2}s}^{l} f(x)v(x)\,dx, & v(l - s) &= v(l) + i\int_{l-s}^{l} g(x)u(x)\,dx.
\end{aligned}
\tag{43}
$$

We now approximate the integrals as follows,

$$
\begin{aligned}
u(0) &= u(\tfrac{1}{2}s) + iv(0)\int_0^{\frac{1}{2}s} f(x)\,dx, & v(0) &= v(s) + iu(\tfrac{1}{2}s)\int_0^{s} g(x)\,dx,\\
u(\tfrac{1}{2}s) &= u(\tfrac{3}{2}s) + iv(s)\int_{\frac{1}{2}s}^{\frac{3}{2}s} f(x)\,dx, & v(s) &= v(2s) + iu(\tfrac{3}{2}s)\int_s^{2s} g(x)\,dx,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
u(l - \tfrac{1}{2}s) &= u(l) + iv(l)\int_{l-\frac{1}{2}s}^{l} f(x)\,dx,\\
& & v(l - s) &= v(l) + iu(l - \tfrac{1}{2}s)\int_{l-s}^{l} g(x)\,dx.
\end{aligned}
\tag{44}
$$

The ratio $u(0)/v(0)$ may be expressed as a continued fraction

$$
\frac{u(0)}{v(0)} = Z_1 + \frac{1}{Y_1 +}\,\frac{1}{Z_2 +}\,\frac{1}{Y_2 +}\cdots + \cfrac{1}{Z_{n+1} + \cfrac{u(l)}{v(l)}},
\tag{45}
$$

where

$$
\begin{aligned}
Z_1 &= i\int_0^{\frac{1}{2}s} f(x)\,dx, & Z_2 &= i\int_{\frac{1}{2}s}^{\frac{3}{2}s} f(x)\,dx, & \cdots, && Z_{n+1} &= i\int_{l-\frac{1}{2}s}^{l} f(x)\,dx,\\
Y_1 &= i\int_0^{s} g(x)\,dx, & Y_2 &= i\int_s^{2s} g(x)\,dx, & \cdots, && Y_n &= i\int_{l-s}^{l} g(x)\,dx.
\end{aligned}
\tag{46}
$$

Knowing this ratio and $K_1 = [f(0)/g(0)]^{1/2}$ and $u(l)/v(l) = [f(l)/g(l)]^{1/2}$, we find the reflection coefficient from (7).

In the case of electromagnetic waves given by (3) we find it convenient to normalize the equations by setting

$$E = u, \qquad \Delta^{-1}(\mu/\epsilon_1)^{1/2}H_t = v, \tag{47}$$

so that

$$f(x) = 2\pi/\lambda_x\,, \qquad g(x) = (2\pi/\lambda_x)\left[1 + \frac{2(n - n_1)}{\Delta^2}\right], \tag{48}$$

where λ_x is given by (16). If we choose $s = \frac{1}{8}\lambda_x$, then

$$Z_1 = Z_{n+1} = i\pi/8, \qquad Z_2 = Z_3 = \cdots = Z_n = i\pi/4,$$

$$Y_m = \frac{i\pi}{4s}\int_{(m-1)s}^{ms}\left[1 + \frac{2(n - n_1)}{\Delta^2}\right]dx. \tag{49}$$

If for $x > l$, the medium is inhomogeneous but such that $K'/2K$ is small, then $u(l)/v(l)$ does not quite equal $[f(l)/g(l)]^{1/2}$. In terms of the reflection coefficient

$$q = e^{2i\vartheta(l)}\int_l^\infty \frac{K'(x)}{2K(x)}\,e^{-2i\vartheta(x)}\,dx, \tag{50}$$

we then find

$$\frac{u(l)}{v(l)} = \frac{1 + q}{1 - q}\,[f(l)/g(l)]^{1/2}. \tag{51}$$

Equation (45) is needed only for that part of the inhomogeneous medium for which $K'(x)/K(x)$ is not small compared with unity.

8. *The Differential Equation for the Reflection Coefficient*

The equivalence between linear differential equations of the second order (or systems of two linear equations of the first order) and the generalized Riccati equation is well known but the physical significance of this equivalence is not. While the wave functions satisfy the former equations, the wave impedance and the reflection coefficient satisfy Riccati equations.[3] Thus equation (7) defines the reflection coefficient at $x = x_1$ in terms of the ratio (the "wave impedance") $Z_i(x_1) = u(x_1)/v(x_1)$ and $K = [f(x_1)/g(x_1)]^{1/2}$. From (4) we find for any x

$$\frac{dZ_i}{dx} = \frac{1}{u}\frac{du}{dx} - \frac{u}{v^2}\frac{dv}{dx} = -if(x) + ig(x)Z_i^2\,. \tag{52}$$

[3]J. R. Pierce, *A note on the transmission line equation in terms of impedance,* Bell System Technical Journal, Volume 22, 1943, pp. 263–265; in this paper equation (52) is obtained directly from physical considerations. L. R. Walker and N. Wax, *Non-uniform transmission lines and reflection coefficients,* Journal of Applied Physics, Volume 17, 1946, pp. 1043–1045; in this paper equation (53) is derived and several applications are given.

This is the differential equation for $Z_i(x)$. From (7) and (52) we obtain

$$\frac{dq}{dx} = 2i\beta(x)q - \frac{1}{2}(1 - q^2)\frac{d \log K}{dx}. \tag{53}$$

If the layer extends from $x = x_1$ to $x = x_2$, then $q(x_2) = 0$. Together with this initial condition, equation (53) determines the reflection coefficient $q(x_1)$.

It is more convenient to express the above equations in terms of the distance from the upper boundary of the layer. This is equivalent to reversing the sign of x in the foregoing equations,

$$\frac{dZ_i}{dx} = if(x) - ig(x)Z_i^2, \qquad \frac{dq}{dx} = -2i\beta(x)q - \frac{1}{2}(1 - q^2)\frac{d \log K}{dx}. \tag{54}$$

Introducing the phase variable $\vartheta(x)$, we have

$$\frac{dZ_i}{d\vartheta} = i\left(K - \frac{Z_i^2}{K}\right), \qquad \frac{dq}{d\vartheta} = -2iq - \frac{1}{2}(1 - q^2)\frac{d \log K}{d\vartheta}. \tag{55}$$

If $| q | \ll 1$, we may neglect q^2 and obtain an approximation for q equivalent to (34). We can then obtain higher order approximations by the perturbation method. Alternatively, we can easily obtain q numerically or on a differential analyzer.

The Theory of Magneto Ionic Triple Splitting

By O. E. H. RYDBECK
Research Laboratory of Electronics,
Chalmers University of Technology,
Gothenburg, Sweden

1. *Introduction and Summary*

Triple splitting of ionospheric waves or rays was reported in 1933 by Eckersley [1]. A few years later, in 1935 and 1936, Toshniwal [2] and Harang [3] reported similar phenomena. However, no serious attempt was made to explain them theoretically. This interesting matter did not attract much attention until after world war II, when the chain of Recording Ionospheric Observatories had been greatly increased.

According to Meek [4] triple splitting was observed in Canada (1943) when the Churchill station commenced operation. Since then the phenomenon was observed at other stations as well. In Hobart, Tasmania, Newstead (1946) observed F_2 triple splitting [5]. Seaton, reporting from College, Alaska in 1947 presents a record of F_2 triple splitting [6] and discusses the possible mechanism of excitation of the third or z-component of the down-coming wave.

At the Kiruna Ionospheric and Radio Wave Propagation Observatory (67° 50′ N, 20° 14. 5′ E), which started regular recording on October 1st 1948, a large number of E- and F-layer triple splits have been recorded [7]. As far as is known to us, E-triple splitting has only been reported earlier by Meek [4] who is also working at high latitudes. The frequent occurrence of ionospheric triple splitting at Kiruna motivated us to attempt a theoretical explanation of the phenomenon. The first outline of the theory, based on the coupled o- and x-wave equations of the author's treatise, "*On the Propagation of Radio Waves*", (Gothenburg 1944), was presented at the annual meeting of the Indian Academy of Sciences in Bombay, December 1949 and was subsequently published [8]. The further theory, including a description of the z-ray paths, was presented at a session of the Societé de Radioélectriciens in Paris, February 1950.

This theory, which forms the main contents of the following communication, proves that for any ionosphere, even a smooth one, a z-wave is excited by the vertically incident ordinary wave in the neighbourhood of and at the ordinary reflection level, at moderately high geomagnetic latitudes. At sufficiently high geomagnetic latitudes this coupling becomes very strong. One can then, for

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

vertical incidence, picture the ionized layer as consisting of coupled *o*- and *z*-transmission lines. Finally, at a critical geomagnetic latitude, for which the inclination of the terrestrial magnetic field, θ_i, is given by

$$\theta_i = \tan^{-1} \{\tfrac{1}{2}([1 + (\omega_H/\nu)^2]^{1/2} - 1)\}^{1/2},$$

where ν is the electronic collisional frequency and ω_H the angular gyrofrequency of the free electrons in the terrestrial magnetic field, the entire *o*-wave is transformed into a *z*-wave and the ordinary trace on the ionospheric record becomes weak. For latitudes higher than the critical one, which for $\nu = 2.10^5$ c/s is about 67°, the *o*-wave rapidly disappears. Only the *z*-and *x*-components remain and the propagation is purely longitudinal. This is in very good agreement with the experimental results.

It is further shown that at low geomagnetic latitudes, where the normal coupling would be far too weak to explain the rare cases of low latitude triple splits, a strong step-like increase in the electron density can produce a *z*-wave of recordable strength, provided this step occurs practically at the regular interaction level and has an extension of about one vacuum wave length. Irregularities in the electron density distribution may therefore, under rare circumstances, be effective in producing triple splits at low latitudes.

2. *General Consideration*

To begin with we assume that the terrestrial magnetic field of strength H lies in the z,y-plane of the rectangular coordinate system with components $H \cdot \cos\theta$ and $H \cdot \sin\theta$ along the respective axes. If z were chosen to be the vertical, $\pi/2 - \theta$ then would denote the inclination of the terrestrial magnetic field, θ_i.

Introducing the average collisional frequency, ν, of the electrons, ω_H = their angular gyrofrequency in the terrestrial magnetic field and ω_c = the critical angular frequency of the ionized medium (of specific density N cm^{-3} of the free electrons) in the absence of the magnetic field, we can write the familiar relations between the x,y,z-components of the electric field strength $\mathbf{E}$, and the polarization, $\mathbf{P}$, in the following form, *viz.* [9, p. 8] (time factor $e^{-i\omega t}$)

$$\begin{aligned} -E_z/4\pi &= x^2(1 + j\delta)\cdot P_z + j\gamma_T\cdot P_x \\ -E_x/4\pi &= x^2(1 + j\delta)\cdot P_x - j\gamma_T\cdot P_z + j\gamma_L\cdot P_y \qquad (1) \\ -E_y/4\pi &= x^2(1 + j\delta)\cdot P_y - j\gamma_L\cdot P_x \end{aligned}$$

where $x^2 = \omega^2/\omega_c^2$, $\gamma_T = x^2\cdot y\cdot \sin\theta$, $\gamma_L = x^2\cdot y\cdot\cos\theta$, $\delta = \nu/\omega$

and $y = \omega_H/\omega$.

We next assume that z is the vertical axis and that the electron density, N, is a function of z (height) only. If we further assume that the incidence of

the (plane) electromagnetic wave is vertical the wave normal will also be vertical, although the direction of maximum energy transport may be more or less oblique (see also Section 6). We therefore put $\partial/\partial y = 0 = \partial/\partial x$. The condition div $\mathbf{D} = 0$, where $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$ is the displacement vector, thus means that $E_z = -4\pi P_z$, as we do not consider constant fields. Under these circumstances we find

$$-E_x/4\pi = \{\alpha + \gamma_T^2/(1 - \alpha)\}\cdot P_x + j\gamma_L \cdot P_y ,$$

$$-E_y/4\pi = \alpha \cdot P_y - j\gamma_L \cdot P_x , \tag{1a}$$

where

$$\alpha = x^2(1 + j\delta).$$

Now for a certain value, u, of the polarization ratio P_y/P_x, the ratios E_x/P_x and E_y/P_y become identical and consequently must belong to the same wave solution. This is very important because it is in a sense the foundation of the following analysis. One immediately finds that two u-values, u_1 and u_2, are possible, *viz.*

$$\begin{matrix} u_1 \\ u_2 \end{matrix} = \frac{j}{-V^2 + j\delta}\{\delta_c \mp (\delta_c^2 + (-V^2 + j\delta)^2)^{1/2}\}, \qquad u_1 \cdot u_2 = 1 \tag{2}$$

where $V^2 = 1/x^2 - 1$ and $\delta_c = \nu_c/\omega$. Here $\nu_c = \omega \cdot y \cdot (\sin^2 \theta)/(2 \cdot \cos \theta)$ is the critical collisional frequency of Appleton-Builder [10].

As V^2 is a function of z (as are also δ_c and δ) the u-ratios are not independent of height. This means, as we will soon see, that the two wave-solutions corresponding to u_1 and u_2 generally are not independent of each other. Since $E_x/P_x = E_y/P_y$, we have

$$E_y^{(1)}/E_x^{(1)} = u_1 , \qquad E_y^{(2)}/E_x^{(2)} = u_2 = 1/u_1 . \tag{3}$$

When $\omega_c = \omega$, $V^2 = 0$, and $u_1 = [\nu_c/\nu - ((\nu_c/\nu)^2 - 1)^{1/2}]$. From this relation one recognizes the old fact that when $\omega_c = \omega$ the polarization angle ϕ ($\tan \phi$ is equal to the ratio between minor and major axes) is zero when $\nu < \nu_c$ (plane polarization) and $\pm \tan^{-1} [(\nu - \nu_c)/(\nu + \nu_c)]^{1/2}$ when $\nu > \nu_c$.

From Maxwell's equations we find [9, p. 12]

$$k_0^2 \cdot D_x + \frac{d^2E_x}{dz^2} = 0,$$

$$k_0^2 \cdot D_y + \frac{d^2E_y}{dz^2} = 0, \tag{4}$$

where $k_0 = 2\pi/\lambda_0$ and λ_0 is the vacuum wave-length.

For the two u-values we further easily find from (1) that

$$D_y^{(1)} = \epsilon_1 \cdot E_y^{(1)}, \qquad D_y^{(2)} = \epsilon_2 \cdot E_y^{(2)}, \qquad D_x^{(1)} = \epsilon_1 \cdot E_x^{(1)}, \qquad D_x^{(2)} = \epsilon_2 \cdot E_x^{(2)}, \tag{5}$$

where ϵ_1 and ϵ_2 are the complex dielectric constants of Appleton-Hartree for

the ordinary and the extraordinary waves. In our notations it is very convenient to write ϵ_1 and ϵ_2 in the following forms, *viz.*

$$\begin{aligned} \frac{\epsilon_1}{\epsilon_2} = \frac{n_1^2}{n_2^2} &= 1 - \frac{(1 + j\delta - B)B}{(1 + j\delta)B - y\cdot\cos\theta[\delta_c \mp (\delta_c^2 + B^2)^{1/2}]} \\ &= \frac{B^2 - y\cdot\cos\theta[\delta_c \mp (\delta_c^2 + B\ )^{1/2}]}{(1 + j\delta)B - y\cdot\cos\theta[\delta_c \mp (\delta_c^2 + B^2)^{1/2}]}, \end{aligned} \tag{6}$$

where $B = -V^2 + j\delta$ and index (1) denotes the ordinary component, etc.

As $D_x = D_x^{(1)} + D_x^{(2)}$ and $D_y = D_y^{(1)} + D_y^{(2)}$, one finds that the formally simple wave equations (4) yield the two coupled wave equations

$$\begin{aligned} \frac{d^2\Pi_1}{dz^2} + (k_0^2\cdot\epsilon_1 - \psi^2)\Pi_1 &= -\Pi_2\cdot\frac{d\psi}{dz} - 2\cdot\frac{d\Pi_2}{dz}\cdot\psi, \\ \frac{d^2\Pi_2}{dz^2} + (k_0^2\cdot\epsilon_2 - \psi^2)\Pi_2 &= -\Pi_1\,\frac{d\psi}{dz} - 2\,\frac{d\Pi_1}{dz}\cdot\psi, \end{aligned} \tag{7}$$

where

$$\Pi_1 = E_x^{(1)}/(1 - u_1^2)^{1/2}, \qquad \text{and} \qquad \Pi_2 = E_y^{(2)}/(1 - u_2^2)^{1/2}. \tag{7a}$$

One further has (with the chosen time-factor)

$$\psi = -\frac{j}{2}\cdot\frac{d(V^2)}{dz}\cdot\frac{\delta_c}{(V^2 - ja_1)(V^2 + ja_2)}, \tag{8}$$

with

$$\begin{aligned} a_1 &= \delta_c + \delta, \\ a_2 &= \delta_c - \delta. \end{aligned} \tag{9}$$

3. *The Unperturbed Wave Functions and the Coupling Coefficient*

For the unperturbed *o*-wave functions we use the very accurate approximations[1]

$$\begin{aligned} \Pi_1^{0(1)} &= \left(\frac{\pi W_1}{2n_1}\right)^{1/2}\cdot e^{j\pi/6}\cdot H_{1/3}^{(2)}(W_1) \\ \Pi_1^{0(2)} &= \left(\frac{\pi W_1}{2n_1}\right)^{1/2}\cdot e^{-j\pi/6}\cdot H_{1/3}^{(1)}(W_1), \end{aligned} \tag{10}$$

[1]Generally very accurate except in the neighborhood of the second zero of ϵ_1. When two zeros come very close (not of immediate interest in this connection) parabolic type wave functions have to be used, see reference [11], p. 23.

where W_1 is the phase integral

$$W_1 = \int_{z_a}^{z} k_0 \cdot n_1 \cdot dz = W_1(z_a\,, z), \tag{11}$$

and z_a the location of the first zero of ϵ_1.

Let us assume that the ionized layer is formed in the height interval $(z_0 - \Delta h) - (z_0 + \Delta h)$, where z_0 denotes the centre of the layer. If further $n_1 = e^{-i\pi}$ at $z_0 - \Delta h$, then $\Pi_1^{0(1)}$ represents a wave incident upon the layer from the $-z$ side (z assumed positive in direction upwards). Neglecting the possible formation of a z-wave the coefficient of reflection at the lower boundary $z_0 - \Delta h$ asymptotically becomes

$$R \sim \exp\{j \cdot 2[W_1(z_a\,, z_0 - \Delta h) - \pi/4]\}, \tag{12}$$

if the layer is not very thin (counted in λ_0) and $\omega \neq \omega_c$.

For the unperturbed z-wave functions we use similar wave functions denoted by $\Pi_2^{0(1)}$, $\Pi_2^{0(2)}$ with

$$W_2 = \int_{z_b}^{z} k_0 \cdot n_2 \cdot dz = W_2(z_b\,, z), \tag{13}$$

where z_b is the position of the appropriate zero. It appears that it is of fundamental importance to study the behaviour of n_1 and n_2 in the z-plane.

We denote by z_{a_0} the interaction level where $\omega_c = \omega$ (also branch point of n_1 when $\nu = 0$) and use the following approximation in the neighbourhood of this level, *viz.*

$$V^2 = \left\{\frac{d}{dz}\left(\frac{\omega_c^2}{\omega^2}\right)\right\}_{z=z_{a_0}} \cdot (z - z_{a_0}) = \gamma_{a_0} \cdot (z - z_{a_0}). \tag{14}$$

Under these circumstances we have

$$\psi = -\frac{j}{2} \cdot \frac{\delta_c/\gamma_{a_0}}{\{z - z_{a_0} - ja_1/\gamma_{a_0}\}\{z - z_{a_0} + ja_2/\gamma_{a_0}\}}, \tag{15}$$

because dV^2/dz varies only slightly through the narrow interaction region. It is extremely interesting from an interaction point of view to find that

$$\int_{-\infty}^{+\infty} \psi\, dz = \begin{cases} 0, & \nu > \nu_c \\ -j\pi/2, & \nu < \nu_c\,. \end{cases} \tag{16}$$

As n_1^2 is zero for $V^2 = j\delta$ we immediately find that the branch point z_a of n_1 is mid-way between the two poles $z_1 = z_{a_0} + ja_1/\gamma_{a_0}$ and $z_2 = z_{a_0} - ja_2/\gamma_{a_0}$ of ψ, *i.e.*

$$\mathcal{I}m\,(z_a) = z_a - z_{a_0} = (a_1 - a_2)/2\gamma_{a_0} = \delta/\gamma_{a_0}\,. \tag{17}$$

It is further important to note that

$$1) \qquad (n_2)_{z=z_a} = 1,$$

and

$$2) \qquad n_1 = n_2 \qquad \text{at the poles.}$$

The two zeros z_b and z_c of n_2^2 are found for $V^2 = y + j\delta$ and $V^2 = -y + j\delta$. They are thus independent of θ.

If we assume a parabolic ionized layer of the form

$$\omega_c^2 = \omega_{cm}^2\{1 - [(z - z_0)/\Delta h]^2\},$$

where ω_{cm} is the angular critical frequency of the layer (for the moment assumed larger than $\omega \cdot (1 + y)$), we have $\mathcal{I}m(z_c) < \mathcal{I}m(z_a) < \mathcal{I}m(z_b)$ as depicted in Figure 1, where the poles of ψ are also shown.

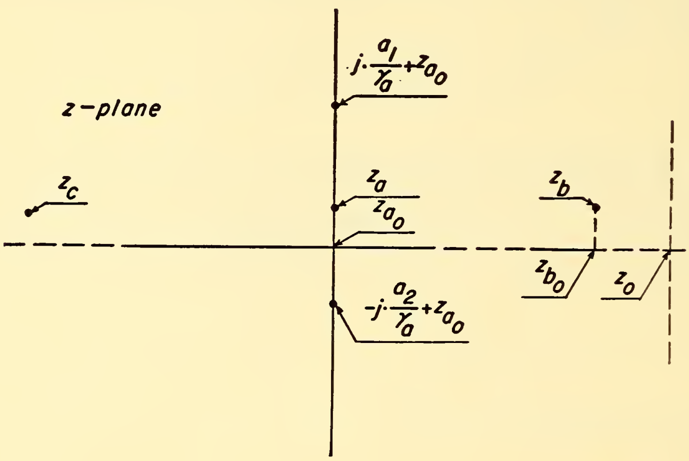


FIG. 1. Location of the zeros of n^2 and the poles of ψ in the z-plane.

Finally it should be noted that $|n_2^2| = \infty$ for $z = z_d$, at which point (branch point of Rai)

$$(\omega_c/\omega)^2_{z=z_d} = \frac{1 + j\delta}{1 + \dfrac{y^2 \cdot \sin^2\theta}{(1 + j\delta)^2 - y^2}}. \tag{18}$$

In Figure 2 we have plotted the variation of ϵ_1 and ϵ_2 for $\nu = 0$, a normal electron density gradient and geomagnetic conditions of the Kiruna Observatory. The unperturbed standing wave functions

$$\Pi_1^{0(s)} = \Pi_1^{0(1)} + \Pi_1^{0(2)} \qquad \text{and} \qquad \Pi_2^{0(s)} = \Pi_2^{0(1)} + \Pi_2^{0(2)},$$

corresponding to these ϵ-variations are shown in Figure 3 for $\nu = 0$, together with $-\mathscr{Im}\ (\psi)$ for $\nu = 0.95 \cdot \nu_c$. This figure demonstrates that the coupling actually takes place within a very narrow region of a few hundred metres effective extension.

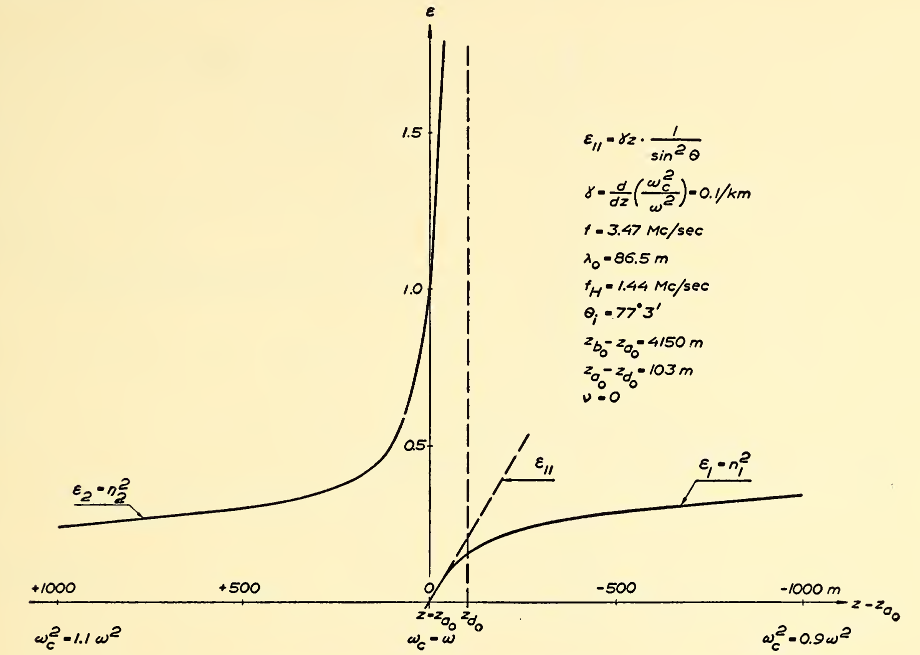


FIG. 2. Index of refraction in the neighbourhood of $\omega_c = \omega$ when $\nu = 0$.

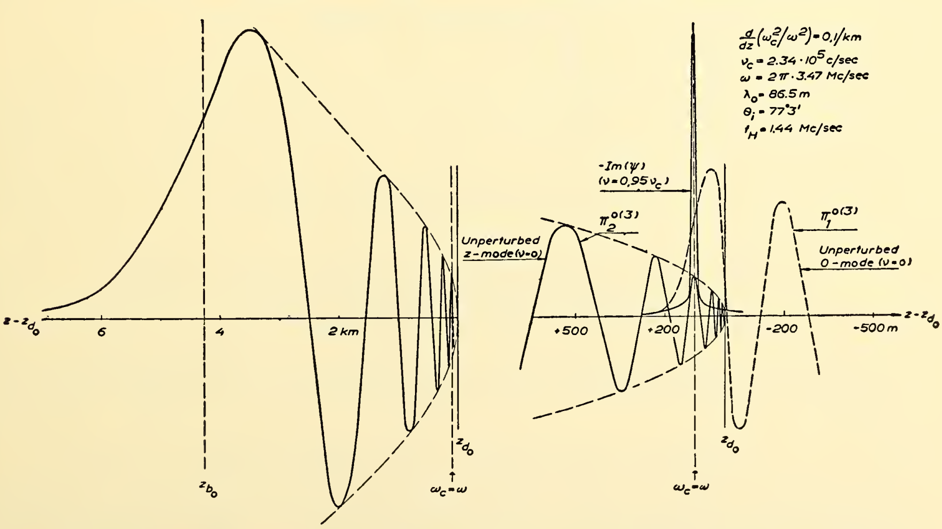


FIG. 3. The coupling coefficient and the unperturbed wave functions in the interaction region.

It is of particular interest in this connection to demonstrate the peculiar nature of ψ for various ν-values. The variation of the important component, $-\mathcal{I}m\ (\psi)$, is therefore shown in Figure 4 from which it is easy to understand that the "interaction integral" (16) is zero when $\nu > \nu_c$.

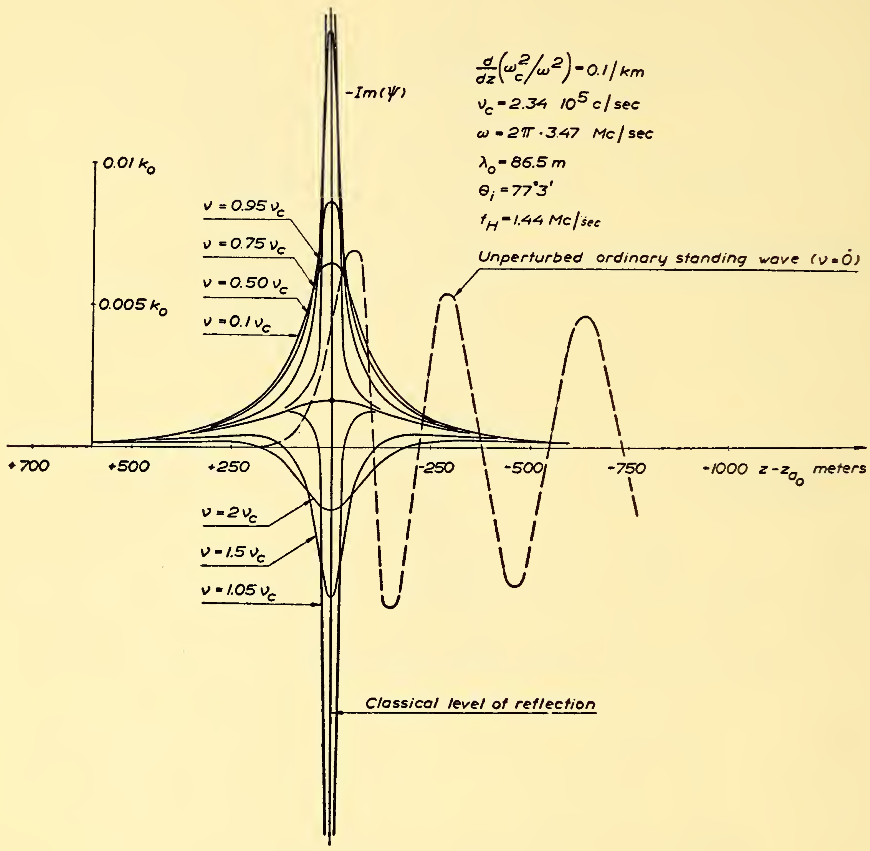


Fig. 4. Demonstrating the nature of $-\mathcal{I}m(\psi)$.

4. *First and Higher Order Approximations of the Z-Wave (Third Component)*

In order to study the first order approximation of the excited z-wave we introduce

$$\xi_1^{(1)}(z) = -\Pi_1^{0^{(1)}} \cdot d\psi/dz - 2\cdot\psi\cdot d\Pi_1^{0^{(1)}}/dz, \text{ etc.} \tag{19}$$

It is necessary to treat the pair $\Pi_1^{0^{(1)}}$, $\Pi_1^{0^{(2)}}$ together, as only the sum of these two wave functions will yield an exponentially decreasing tail along the positive

real axis (z real and $>z_{a_0}$). The first order z-wave propagating in positive direction thus becomes

$$\Pi_2^{1^{(1)}} = -\Pi_2^{0^{(1)}} \cdot \frac{1}{w_2^0} \int_{b_1}^{z} \Pi_2^{0^{(2)}} \cdot (\xi_1^{(1)} + \xi_1^{(2)})\, dz, \tag{20}$$

where w_2^0 is the Wronskian of the pair $\Pi_2^{0^{(1)}}$, $\Pi_2^{0^{(2)}}$. With our form of functions $w_2^0 = j \cdot 2k_0$. Similarly for the wave running in opposite direction we obtain

$$\Pi_2^{1^{(2)}} = \Pi_2^{0^{(2)}} \cdot \frac{1}{w_2^0} \int_{b_2}^{z} \Pi_2^{0^{(1)}} \cdot (\xi_1^{(1)} + \xi_1^{(2)})\, dz. \tag{21}$$

In these formulae b_1 and b_2 denote lower and upper boundaries of the effective interaction region.

It is convenient to write these integrals the following way, *viz.*

$$\begin{aligned} \Pi_2^{1^{(1)}} &= -\Pi_2^{0^{(1)}} \cdot \frac{1}{w_2^0} \cdot \int_{b_1}^{z} \{\Pi_1^{0^{(2)}} \cdot \Pi_2^{0^{(2)}} \cdot (\Pi_1^{0^{(2)}\prime}/\Pi_1^{0^{(2)}} - \Pi_2^{0^{(2)}\prime}/\Pi_2^{0^{(2)}})\psi \\ &\qquad + \Pi_1^{0^{(1)}} \cdot \Pi_2^{0^{(2)}} \cdot (\Pi_1^{0^{(1)}\prime}/\Pi_1^{0^{(1)}} - \Pi_2^{0^{(2)}\prime}/\Pi_2^{0^{(2)}})\psi\}\, dz \\ &= -\Pi_2^{0^{(1)}} \cdot \frac{1}{w_2^0} \cdot \int_{b_1}^{z} f_1(z) \cdot \psi \cdot dz, \end{aligned} \tag{22}$$

and similarly

$$\begin{aligned} \Pi_2^{1^{(2)}} &= \Pi_2^{0^{(2)}} \cdot \frac{1}{w_2^0} \cdot \int_{b_2}^{z} \{\Pi_1^{0^{(2)}} \cdot \Pi_2^{0^{(1)}} \cdot (\Pi_1^{0^{(2)}\prime}/\Pi_1^{0^{(2)}} - \Pi_2^{0^{(1)}\prime}/\Pi_2^{0^{(1)}})\psi \\ &\qquad + \Pi_1^{0^{(1)}} \Pi_2^{0^{(1)}} \cdot (\Pi_1^{0^{(1)}\prime}/\Pi_1^{0^{(1)}} - \Pi_2^{0^{(1)}\prime}/\Pi_2^{0^{(1)}})\psi\}\, dz. \end{aligned} \tag{23}$$

For the moment we are principally interested in equation (22) which yields the first order approximation of the wave propagating towards z_b. As far as the interaction integral is concerned, replacing b_1 and b_2 by $-\infty$ and $+\infty$ should not appreciably affect its value if z does not lie between b_1 and b_2 and these boundaries are well outside the main interaction zone.

If we now deform the path of integration to follow mainly the straight line $\mathcal{I}m\,(V^2) = -j(a_2 + \epsilon) = -j\{(\delta_c - \delta) + \epsilon\}$, where ϵ is supposed to be a very small quantity, it is possible to show that the main contribution to the interaction integral comes from the residue at the pole $-ja_2$ of ψ. As a matter of fact, the remaining contribution (from the rest of the path) is of the order of the second term in the asymptotic expansion of the dominant wave function at the pole. When therefore $\nu > \nu_c$, $\Pi_2^{1^{(1)}} = 0$.

In spite of the fact that the poles of ψ are also branch points of n_1 and n_2 (where the cuts joining their Riemann surfaces begin) we can use our wave function approximations with good accuracy at these points when the coupling is small ($|\psi|$ small along axis of reals). This can be demonstrated approxi-

mately in the following manner. Introducing $n_1 = [1 + l_1(z)]^{1/2}$, we know that $\Pi_1^{(2)}$, $\Pi_1^{(1)}$ are exact solutions of the wave equation

$$d^2\Pi/dz^2 + (k_0^2 \cdot n_1^2 - Q)\Pi = 0,$$

where

$$Q = -l''(z)/4\{1 + l(z)\} + 5/16[l'(z)/\{1 + l(z)\}]^2 - 5/4(W_1'/3W_1)^2.$$

As $(\delta_c^2 + B^2)^{1/2} = -[(V^2)' \cdot \delta_c/2j\psi]^{1/2}$ (appearing in $l(z)$; see equation (6)), one finds that Q contains terms of type $\psi^{1/2}, \psi, \psi^{3/2}$. To add these terms to the original wave equations should hardly affect the result when ψ is small. The approximation is slightly rougher than the neglect of the ψ^2 term actually present in the wave equations.

The distance between the poles and z_a is quite large. For the locality in question (Kiruna) $\nu_c \cong 2.1 \cdot 10^5$ sec^{-1}, and with the electron density gradient chosen this distance is greater than 100 metres. The amplitude of W_1 at the poles therefore will be of the order 3. The corresponding phase angle, with the chosen orientation of the n-planes, lies between $-\pi/2$ and $-3\pi/4$ at z_1 and between $\pi/2$ and $3\pi/4$ at z_2. For W_2 the amplitude is several times as large with phase angles approximately $\mp\pi/2$.

It is therefore possible to use the asymptotic representations of the Hankel-functions at the poles. As $n_1 = n_2$ at the poles, one infers that the first term of (22) practically disappears. Virtually only the incident wave function produces a z-wave into the layer. It should be borne in mind, namely, that the reflected wave function, $\Pi_1^{0^{(2)}}$ must be present to make the integral practically zero over the rest of the path. Furthermore, in a region of rapidly changing refractive index, only a special treatment of $\Pi_1^{0^{(1)}}$ and $\Pi_1^{0^{(2)}}$ will disclose how much of them is to be regarded as momentary waves in both directions [11].

At the poles

$$[\Pi_1^{0^{(1)}\prime}/\Pi_1^{0^{(1)}} - \Pi_2^{0^{(2)}\prime}/\Pi_2^{0^{(2)}}] \sim 2jk_0 \cdot n_{pole},$$

and therefore

$$\text{(24)} \quad \Pi_2^{1^{(1)}}(z) \cong \Pi_2^{0^{(1)}}(z) \cdot \exp\{-j[W_1(z_a, z_2) - W_2(z_b, z_2)]\} \cdot e^{-j\pi/2} \cdot \pi/2, \quad z > b_2.$$

We should have started with unit amplitude of the incident wave at $z_0 - \Delta h$ (the layer boundary). The $\Pi_1^{0^{(1)}}(z)$-function of the interaction integral should therefore be replaced by $\Pi_1^{0^{(1)}}(z)/\Pi_1^{0^{(1)}}(z_0 - \Delta h)$. If $z > b_2$, but $z \neq z_b$ this yields the useful result

$$\text{(24a)} \quad \begin{aligned} \Pi_2^{1^{(1)}}(z) &\cong \left(\frac{1}{n_2}\right)^{1/2} \cdot \exp\{j[W_1(z_{a_0}, z_0 - \Delta h) + W_2(z, z_{a_0})]\} \\ &\quad \cdot \frac{\pi}{2} \exp\left\{j\left[W_2(z_{a_0}, z_2) - W_1(z_{a_0}, z_2) - \frac{\pi}{2}\right]\right\}. \end{aligned}$$

As the first exponential is the straight phase integral up to z (though with

exchanged refractive indices after the z-wave "takes over") we are led to introduce the first order transmission coefficient $T_+^{(1)}$ of direction into the layer, *viz.*

$$(25)\qquad \begin{aligned} T_+^{(1)} &\cong \frac{\pi}{2} \exp\left\{j\left[W_2(z_{a_0}, z_2) - W_1(z_{a_0}, z_2) - \frac{\pi}{2}\right]\right\}, && \nu < \nu_c \\ T_+^{(1)} &\cong 0, && \nu > \nu_c . \end{aligned}$$

As $z_2 - z_{a_0} \cong -j(\nu_c - \nu)/\omega \cdot \gamma_{a_0}$, we see that $|T_+^{(1)}|$ approaches $\pi/2$ as ν approaches ν_c. This indicates a very strong coupling as indeed one would expect. When ν approaches ν_c the mathematical reflection point, z_a, comes farther and farther away from the physical interaction point z_{a_0}. At the same time the interaction pole, z_2, comes closer and closer to the physical interaction point and more and more of the incident wave energy is transformed into a z-wave. Finally, at the point of "resonance" ($\nu = \nu_c$) there appears to be practically complete transformation. When the coupling becomes so strong, however, our first order approximations cannot be accurate enough and it is surprising that they are as good as they are. In (25) one would only expect a coefficient of about 1 instead of $\pi/2$ when ν approaches ν_c. However, it is not too complicated to obtain approximate higher order approximations if we make use of the fact that

$$(26)\qquad \Pi_1^{1(1)} = \frac{\Pi_1^{0(1)}}{\overset{0}{\omega_2} \cdot \overset{0}{\omega_1}} \cdot \int_{b_1}^{z} f_2(s)\psi(s)\, ds \cdot \frac{1}{2} \int_{b_1}^{z} f_1(r)\psi(r)\, dr,$$

etc., where $\Pi_1^{1(1)}$ denotes the first order correction in $\Pi_1^{0(1)}$.

Introducing $T_+^{(1)} = \pi/2 \cdot \exp\{-j\pi/2 - \tau\}$, we find the following approximate total transmission coefficient

$$(27)\qquad \begin{aligned} T_+ &\cong \frac{\pi}{2} \cdot e^{-j\pi/2} \cdot \frac{e^{-\tau}}{1 + \frac{\pi^2}{16} \cdot e^{-2\tau}}, && \nu < \nu_c \\ T_+ &\cong 0, && \nu > \nu_c . \end{aligned}$$

For $\nu = \nu_c$ this yields $(T_+)_{\nu=\nu_c} \cong \pi/2/(1 + \pi^2/16) \cong 0.97$. For $|e^{-\tau}| = 1/2$ (the imaginary part of τ is quite small as will soon be shown) similarly $|T_+| \cong 0.72$.

As n_1 and n_2 are representations of the same function in different, interconnected Riemann-surfaces only (these properties have been discussed in an interesting recent paper by T. L. Eckersley [12] it might be of interest to search for a connection formula between the wave functions in the principal n_1 and n_2-planes. A complete connection would yield as result a formula containing all waves running up and down and between the various branch points. Such general connections, which are very difficult, are not of immediate technical interest as many of the components are highly attenuated and it is therefore inadequate to try to connect the waves reflected at z_a and z_b.

If the z-wave functions $\Pi_2^{0^{(1)}}$, $\Pi_2^{0^{(2)}}$ are "carried" through the cut round z_2 back to the axis of reals, but now in the o-plane, one immediately infers from the asymptotic expansions of these functions that they transfer to o-waves

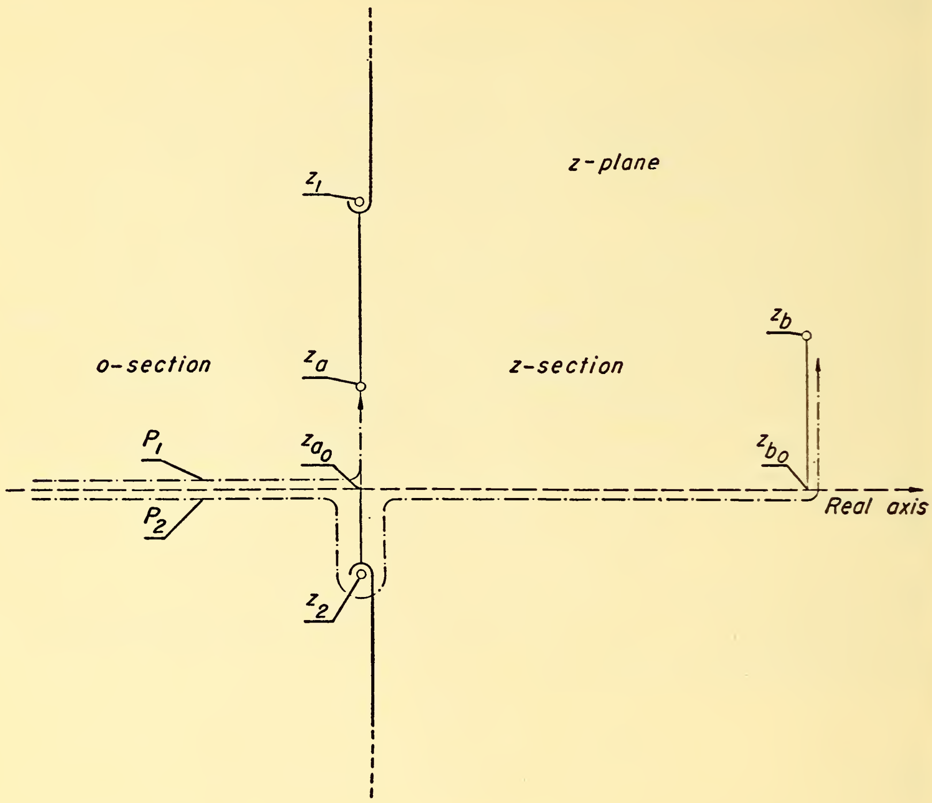


FIG. 5. Transition cuts between two of the Riemann-surfaces.

along the axis of reals at a sufficient distance from z_{a_0}. The transformation coefficient apparently gets the "geometrical" value

$$\exp \{\mp j[W_2(z_{a_0}, z_2) - W_1(z_{a_0}, z_2)]\} = e^{\pm\tau}.$$

Applying the methods of connection already used by the present author [11] we connect the two function groups $\Pi_1^{0^{(1)}} + A\cdot\Pi_1^{0^{(2)}}$ and $B(\Pi_2^{0^{(1)}} + \Pi_2^{0^{(2)}})$. The effective reflection coefficient then becomes $R_{eff} = A\cdot R$, where R is the regular reflection coefficient (12) when the z-waves are not considered. One finds when $\nu < \nu_c$ that

$$(28) \qquad R_{eff} \cong \exp \{2j[W_1(z_a, z) - \pi/4]\}(1 + u^2/2)/(1 - u^2/2), \qquad \nu < \nu_c$$

where

$$u^2 = e^{-2\tau} \exp \{2j[W_2(z_b\,, z_{a_0}) - W_1(z_a\,, z_{a_0})]\}.$$

As $|u^2| \ll 1$ we have approximately

$$R_{\text{eff}} \cong \underbrace{\exp \{2j[W_1(z_a\,, z) - \pi/4]\}}_{R_0}$$

(28a)

$$+ \underbrace{e^{-2\tau} \cdot \exp \{2j[W_2(z_b\,, z_{a_0}) + W_1(z_{a_0}\,, z)]\}}_{R_z}.$$

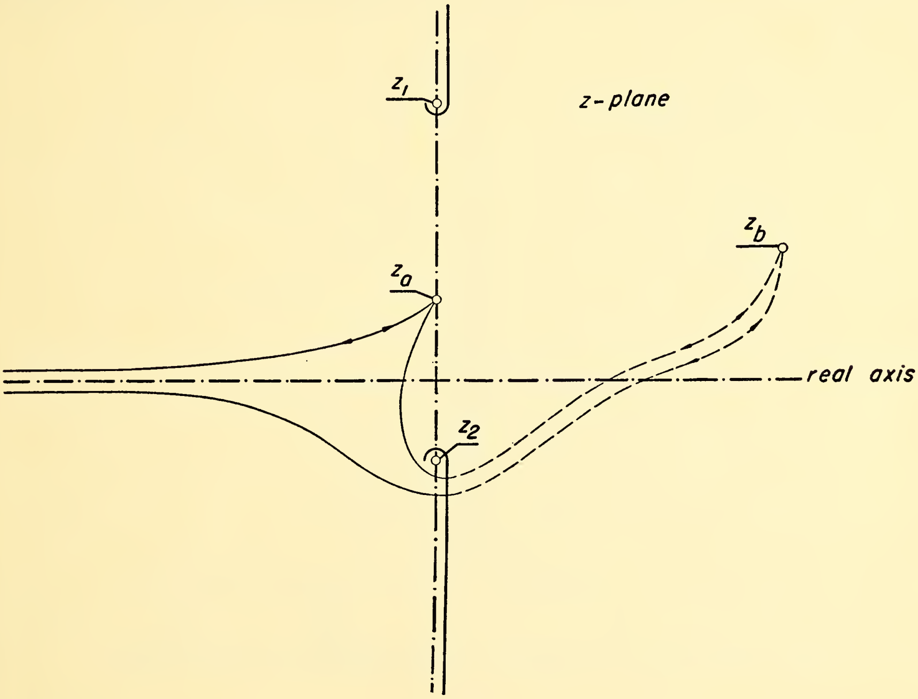


FIG. 6. Depicting complex of waves contained in the effective reflection coefficient.

Expansion (28) shows that it is possible to consider the branch-points as having local "geometrical" reflection coefficients $e^{\pm j\pi/2}$ and transmission coefficients $2^{1/2} \cdot e^{\pm j\pi/4}$. The "geometrical" attenuation from branch-point to branch-point becomes $d = \ln 2 + |\mathcal{I}m \{W_2(z_{a_0}\,, z_2) - W_1(z_{a_0}\,, z_2)\}|$. It must be borne in mind that this representation is a mere construction. However, it is possible to give these "constructed" coefficients a simple physical interpretation [11, p. 17].

When $\nu > \nu_c$, the real axis passes straight through the cut into the z-section and the o-wave automatically is "deformed" into a z-wave, *i.e.* we have

longitudinal transmission. This simply explains why $T_+ = 0$, when $\nu > \nu_c$. The connection relations now yield

$$u^2 = \exp\{2jW_1(z_b, z)\}, \qquad \nu > \nu_c$$

and thus approximately

$$(29) \quad R_{\text{eff}} \cong \underbrace{\exp\{2j[W_1(z_a, z) - \pi/4]\}}_{R_0} + \underbrace{\exp\{2j[W_1(z_b, z) - \pi/4]\}}_{R_z}, \quad \nu > \nu_c .$$

Of these terms the first one represents that which still remains reflected from around the 0-level of reflection. This term disappears fairly rapidly but still

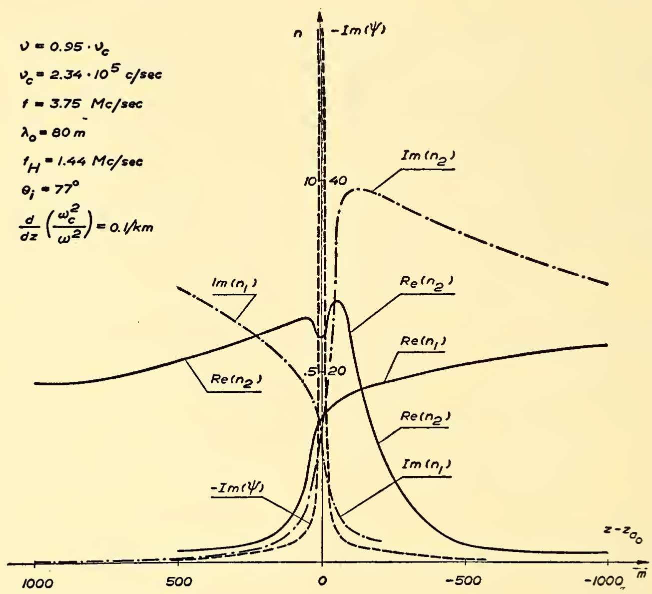


FIG. 7. Depicting the variation of n_1 and n_2 through the coupling range.

exists when $\nu > \nu_c$. The triple split therefore does not abruptly change to a double split type z, x when $\nu \geq \nu_c$. This is also clearly shown in Figures 7, 8 and 9 which depict the real and imaginary components of n_1 and n_2 as functions of $z - z_{a_0}$ along the real axis through the coupling range.

When $\nu = \nu_c$ the transition from the n_1- to the n_2-branch is clearly shown. This transition makes $dn_1/dz = \infty = dn_2/dz$ when $z = z_{a_0}$. This produces a considerable part of the 0-level reflection still present when $\nu = \nu_c$. When $\nu = 1.05 \cdot \nu_c$, Figure 9 shows that there remains a fairly noticeable change in n_1 around the z_{a_0}-level. This produces the reflection mathematically represented by the first term in (29).

Finally, from (27), (28) and (29) it appears that with respect to the com-

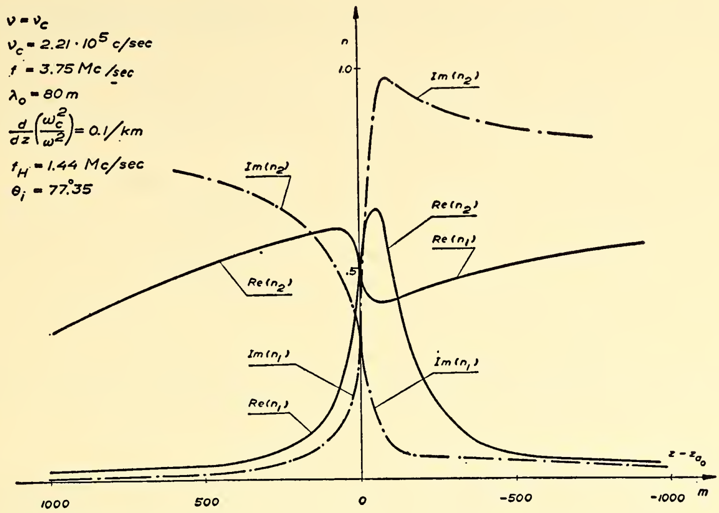


FIG. 8. Depicting the variation of n_1 and n_2 through the coupling range.

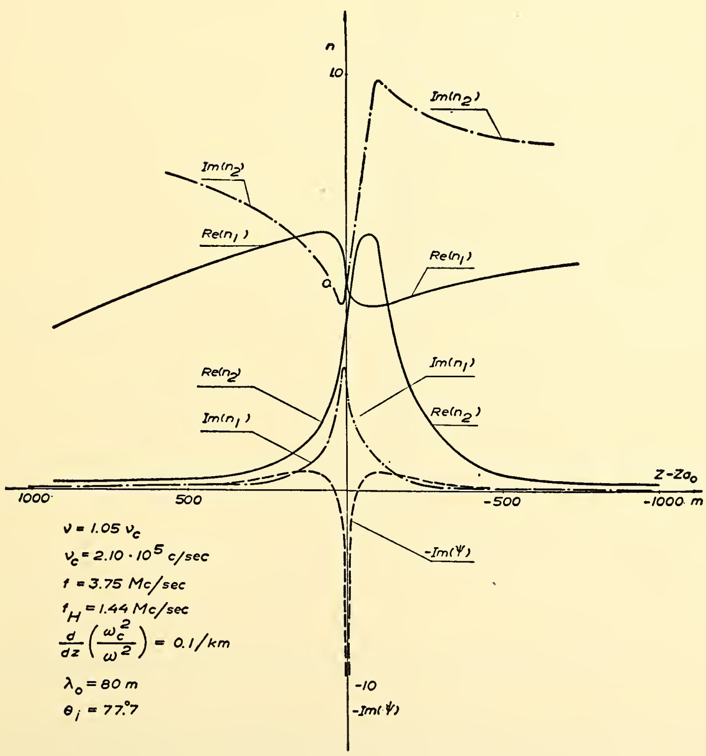


FIG. 9. Depicting the variation of n_1 and n_2 through the coupling range.

ponent reflected from the z-level (*i.e.* at z_b) the effective transmission coefficient becomes

$$
(30)\qquad T_{0z} \cong \begin{cases} \dfrac{\pi}{2}\cdot e^{-i\pi/2}\cdot\dfrac{e^{-\tau}}{1+\dfrac{\pi^2}{16}\cdot e^{-2\tau}}, & \text{(one finds } T_- = -T_+) \\[2ex] e^{-\tau}, & \left.\begin{array}{l}(\nu \text{ small})\\ (\nu \text{ large})\end{array}\right\} \nu < \nu_c \\[2ex] 1, & \nu > \nu_c\,. \end{cases}
$$

For all practical applications it is sufficient to use the simple form

$$
T_{0z} = e^{-\tau}, \qquad \nu \leq \nu_c\,; \qquad T_{0z} = 1, \qquad \nu \geq \nu_c\,.
$$

The fact that the coupling between the wave function disappears when $\nu > \nu_c$ is natural, as has already been stated, because we have longitudinal

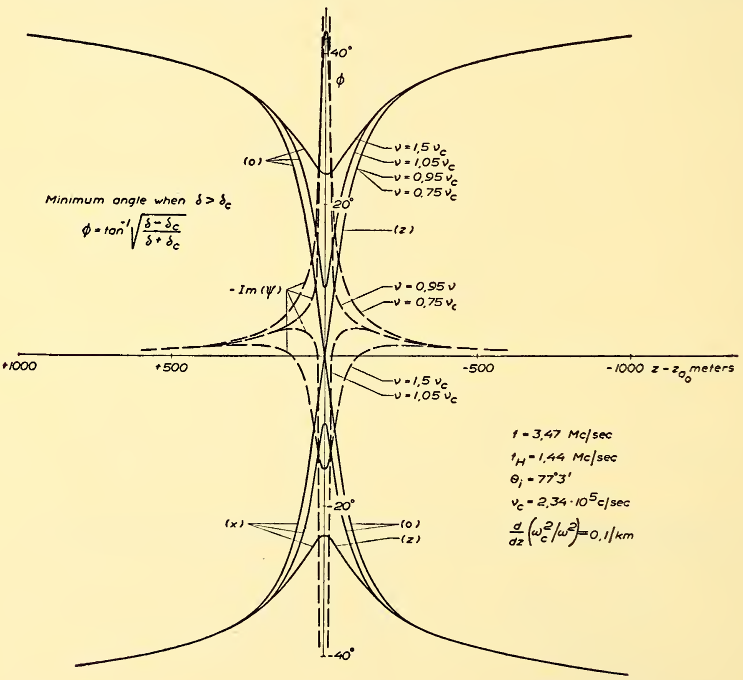


FIG. 10. Depicting the variation of the polarization angle through the coupling region.

transmission when $\nu > \nu_c$. However, if we take a look at Figure 10 depicting the variation of the polarization angles through the coupling region, we will find another way of explaining the phenomenon. Only when $\nu < \nu_c$ do the

polarization curves cross each other, which appears at z_{a_0}. It is easy to understand that coupling is much more likely to take place between the waves when the polarization is the same than when it is widely different. Even if coupling were present, one infers from the shape of the n_2-curves when $\nu > \nu_c$ (Figure 9) that no wave of any significance could be excited or propagate.

5. *The Transmission Coefficient as a Function of* ν

In accordance with (30) we have

$$T_{0z}^{\mp} \cong \exp\left\{j\left[k_0 \int_{z_{a_0}}^{z_2} (n_2 - n_1)\, dz \pm \pi/2\right]\right\}, \qquad \nu \leq \nu_c\,. \tag{30a}$$

Introducing $-V^2 + j\delta = j\mu\delta_c$, we obtain the following simple expressions

$$\begin{matrix} n_1^2 \\ \\ n_2^2 \end{matrix} = \frac{1 \mp (1 - \mu^2)^{1/2} + (1/2)\mu^2 \cdot \tan^2 \theta}{1 \mp (1 - \mu^2)^{1/2} - j\mu(1 + j\delta)/\cos \theta}, \tag{31}$$

where μ varies from -1 to $+1$, when z varies from z to z_2 (along the imaginary axis). In Figure 11 we have plotted the variation of n_1 and n_2 along the pole axis.

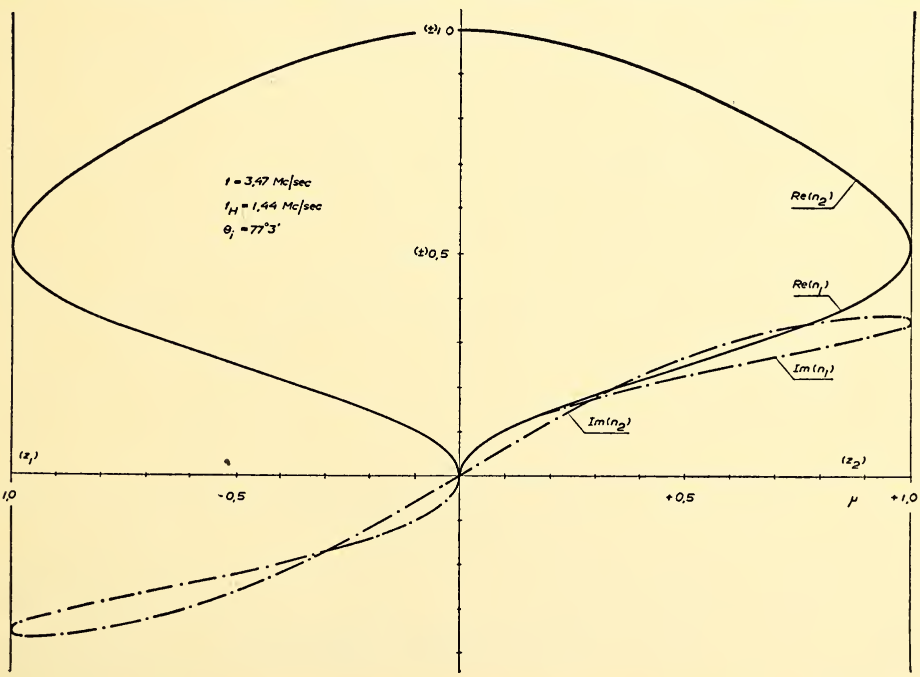


Fig. 11. Showing the variation of refractive indices along the pole axis.

axis. We note the small difference between $\mathscr{I}m\,(n_2)$ and $\mathscr{I}m\,(n_1)$, which indicates that the real part of the phase integral in (30a) is very small.

With $\mu = \nu/\nu_c$ we can write

(30b) $$T_{0z}^{\mp} \cong \exp \{j[k_0\delta_c(1 - \mu)M(\mu)/\gamma_{a_0} \pm \pi/2] - k_0\delta_c(1 - \mu)N(\mu)/\gamma_{a_0}\},$$

where

(32) $$M(\mu) = \frac{1}{1 - \mu}\int_\mu^1 \{\mathscr{I}m\,(n_2) - \mathscr{I}m\,(n_1)\}\,d\mu,$$

and

(33) $$N(\mu) = \frac{1}{1 - \mu}\int_\mu^1 \{\mathscr{R}e\,(n_2) - \mathscr{R}e\,(n_1)\}\,d\mu.$$

In accordance with Figure 12, which depicts the variation of the mean values of refractive index differences (for the Kiruna values used throughout) $M(\mu)$ can be safely neglected, being of the order of 0.05 only.

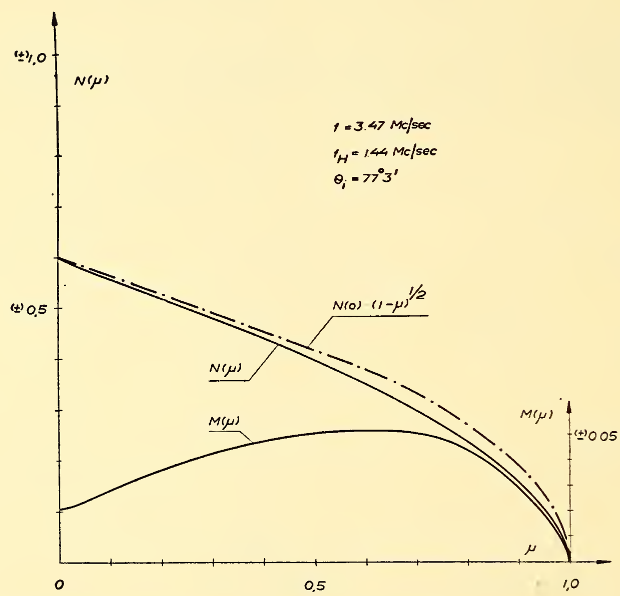


FIG. 12. Depicting the mean values of refractive index differences as functions of μ along the pole axis.

Therefore with good accuracy

(31a) $$T_{0z}^{+}\cdot T_{0z}^{-} \cong \exp\{-2k_0\delta_c(1 - \mu)N(\mu)/\gamma_{a_0}\}.$$

From (31a) we finally obtain for the parabolic layer (of half thickness Δh) the practically useful formula

(34) $$T_{0z}^{+}\cdot T_{0z}^{-} \cong \exp\{-\nu_c\Delta h(1 - \nu/\nu_c)N(\nu/\nu_c)(\omega/\omega_{c_m})/[\omega_{c_m}^2/\omega^2 - 1]^{1/2}\}, \quad \omega < \omega_{c_m}$$

where c_0 is the velocity of light in vacuum. At least for Kiruna conditions $N(\mu)$ according to Figure 12 is well represented by $N(0)(1 - \mu)^{1/2}$ for practical purposes.[2] We therefore obtain the very useful practical approximation of (34), *viz.*

$$(35) \qquad | T_{0z} |^2 \cong \exp \{-\nu_c \Delta h(1 - \nu/\nu_c)^{3/2} \cdot N(0)(\omega/\omega_{cm})/[\omega_{cm}^2/\omega^2 - 1]^{1/2}\}.$$

It is apparent from Figure 12, relations (34) and (35), that

$$(d \mid T_{0z} \mid^2/dz)_{\nu=\nu_c} = 0.$$

Figure 13 illustrates the result (34) for typical E-layer values and Kiruna geomagnetic conditions.

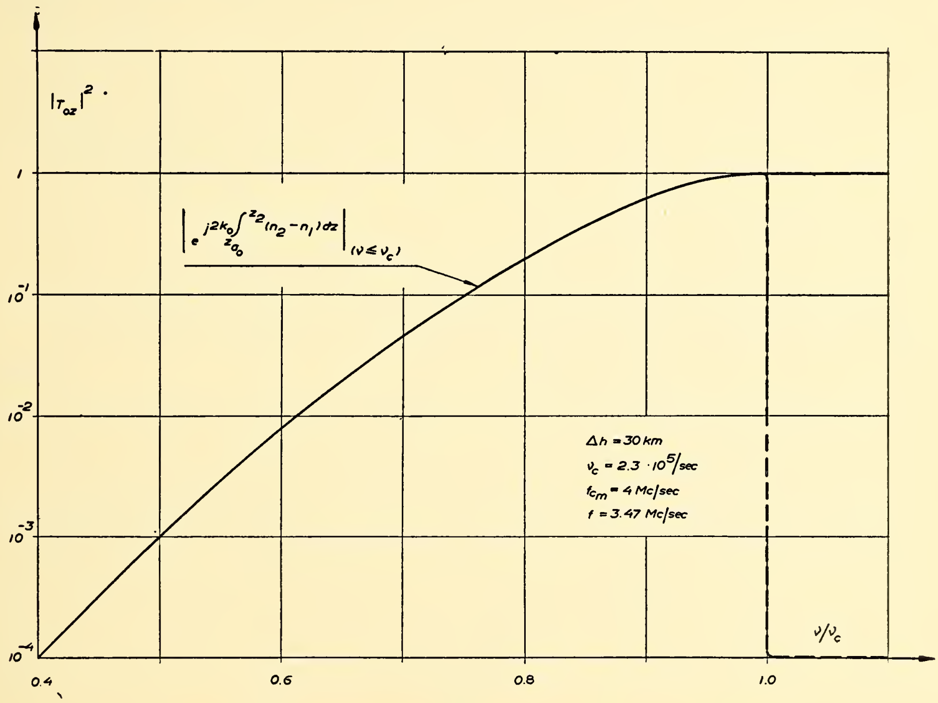


FIG. 13. Total transmission coefficient as a function of ν.

In Figure 14 we have plotted the approximate variation of ν_c with geomagnetic latitude. For comparison the ionospheric height corresponding to the ν-scale has also been included. First, when we come to a geomagnetic latitude of say 60° (the approximate geomagnetic latitude of Oslo) does the longitudinal propagation develop in the lower E-layer for vertical incidence. For low, high and very high geomagnetic latitudes the ionospheric virtual height

[2]The possibilities of approximating $N(\mu, \theta)$ by $N(0, \theta)(1 - \mu)^{1/2}$ over a wider range of θ values will be investigated shortly.

traces, as functions of wave frequency, will therefore, in accordance with the theory presented, appear as shown in Figure 15. This is in good agreement with experimental results.

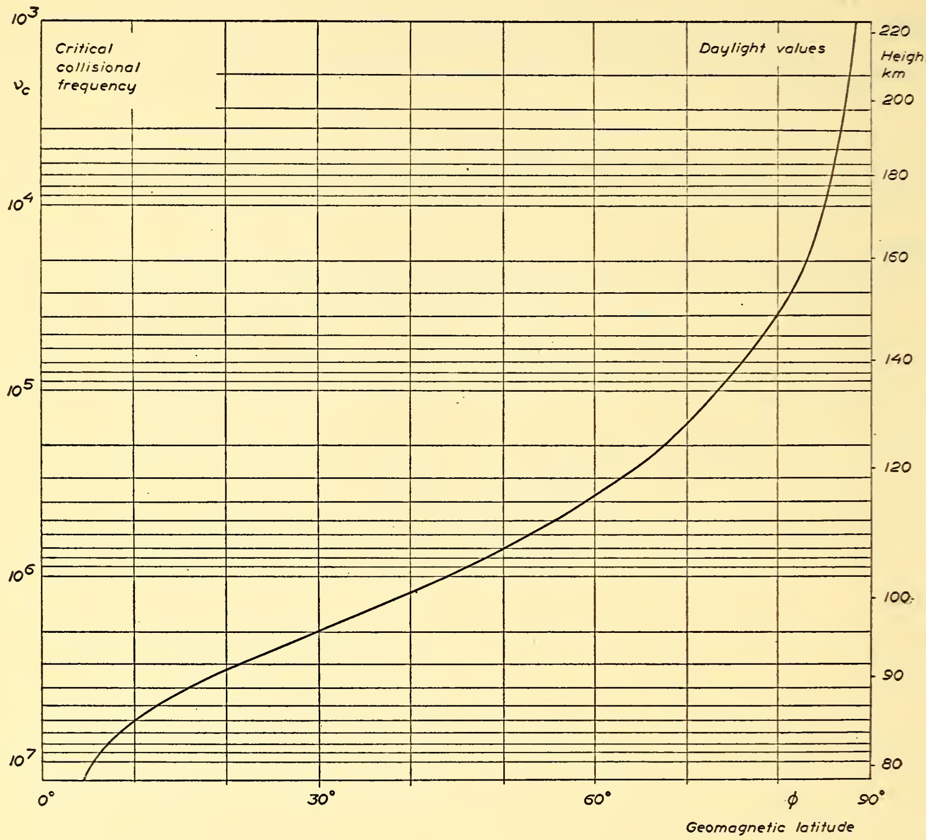


FIGURE 14

At the Kiruna Observatory, where $\nu_c \cong 2.1 \cdot 10^5$ c/s, triple splitting is observed very frequently in the central and upper regions of the E-layer. It is observed regularly in the early mornings and late afternoons when the ionosphere is quiet or only moderately disturbed. The frequency of F_2 triple splits is smaller. At lower geomagnetic latitudes (southern Scandinavia) the situation generally is reversed. Therefore, let us for a moment study the ratio R_z/R_0. Assuming that for $\nu < \nu_c$

$$| \mathcal{I}m \{W_2(z_b , z_{b_0})\} - \mathcal{I}m \{W_1(z_a , z_{a_0})\} | \ll | \mathcal{I}m \{W_2(z_{b_0} , z_{a_0})\} |$$

we have

$$|R_z/R_0| \cong \exp\left\{-2k_0\left[\left|\int_{z_{a_0}}^{z_{b_0}} \mathcal{I}m\,(n_2)\,dz\right| \right.\right. \tag{36}$$
$$\left.\left. + \,\delta_c(1-\mu)^{3/2}\cdot N(0)/\gamma_{a_0}\right]\right\} = e^{-2a}, \qquad \nu < \nu_c\,.$$

Further assuming that (as far as order of magnitude is concerned) for higher latitudes

$$\left|\int_{z_{a_0}}^{z_{b_0}} \mathcal{I}m\,(n_2)\,dz\right| \approx (z_{b_0} - z_{a_0})\,\delta\kappa_1/(1+y),$$

where κ_1 may be smaller or larger than one, and noting that at least approxi-

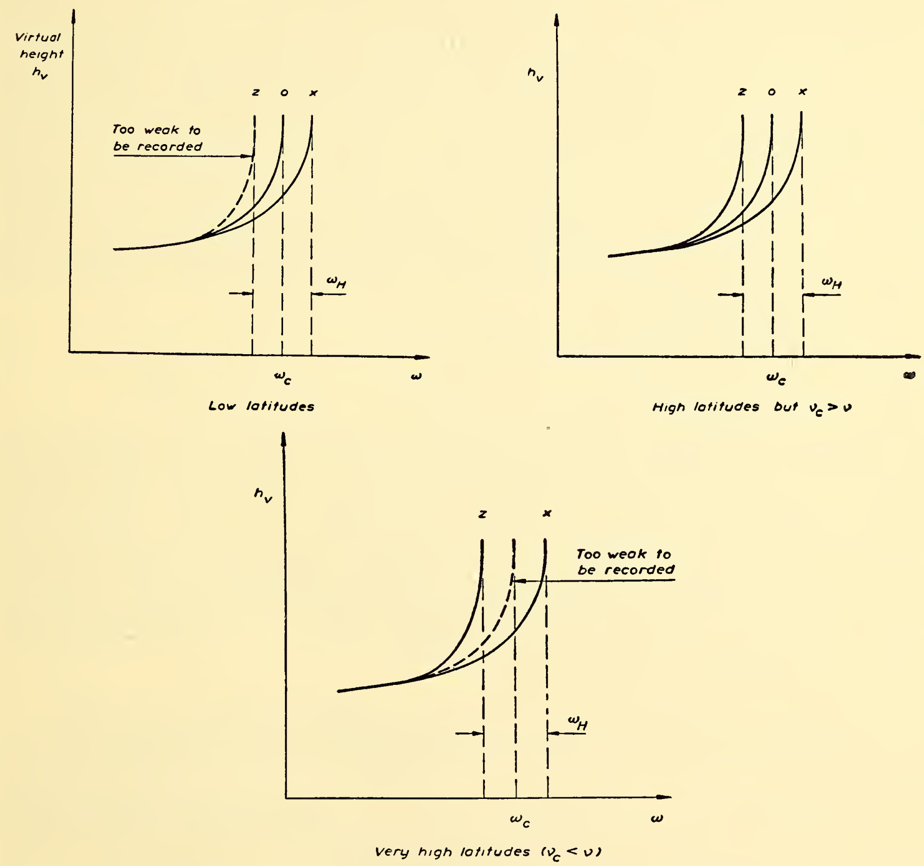


FIG. 15. Depicting ionospheric virtual heights as functions of wave frequency for low, high and very high geomagnetic latitudes.

mately $(z_{b_0} - z_{a_0})d(\omega_c^2/\omega^2)/dz \cong \kappa_2 \cdot y$, where κ_2 probably may be somewhat larger than one, we find

$$q \approx \underbrace{\{N(0)\cdot(1-\mu)^{3/2} + \mu\kappa_1\kappa_2\cdot y/(1+y)\}}_{\chi} \nu_c/c_0\cdot\gamma_{a_0}\,. \tag{37}$$

A normal value of $N(0)$ is about 0.5 (for the Kiruna conditions chosen actually 0.6, see Figure 12). For E-layer conditions $\kappa_1\cdot\kappa_2 y/(1+y)$ may be about the same and for the F-layer considerably smaller. This means that for the E-layer χ may be only slightly varying whereas it should decrease with increasing μ, ν/ν_c, for the F-region. This means that in the F-region one is likely to record low level triple splits more frequently than at high levels. This is in substantial agreement with our Kiruna results.

A comparison between the E- and F-layers depends very much upon the respective gradients γ_{a_0}. At Kiruna γ_{a_0} often is so large for the auroral E-layer that q becomes quite small. At lower latitudes this is not so and as χ moreover is smaller for the F-layer, F-triple splits may be recorded more frequently. This is in good agreement with results in southern Scandinavia. In all cases, places with low ν_c (see relation (37)), *i.e.* at very high latitudes, are the places where strong triple splits are likely to occur. However, triple splits have been reported from places where ν_c is so large as to make q too great for any triple splits to be recorded. One therefore naturally asks the question if a sudden change in electron density, almost a discontinuity, might not produce the desired $0 - z$ coupling at the lower latitudes. A brief investigation will show that this is not generally so.

Let us assume that V^2 has the form

$$V^2 = \gamma_{a_0}\cdot z + \beta\cdot\tanh\{(z-\epsilon)/x_0\}, \qquad \beta,\ \epsilon,\ \text{and } x_0 \text{ real.} \tag{38}$$

The smaller x_0 becomes, the more step-like becomes the electron density distribution. If $(\epsilon/x_0)^2 \ll 1$, the step function can be considered as lying within the main interaction range. If, on the other hand, $(\epsilon/x_0)^2 \gg 1$ the step lies outside the interaction range. Further, it should be noted that the poles of V^2, $z = \epsilon + jx_0(\pi/2 \pm n\cdot\pi)$, are not poles of ψ. Therefore, as far as the production of the z-trace is concerned, the step function has very little influence when $(\epsilon/x_0)^2 \gg 1$ (except that the interaction level $V^2 = 0$ is somewhat displaced). Writing $z = x + jy$ we have

$$\begin{aligned} \mathscr{I}m\,(V^2) &= \alpha\cdot y + \beta\cdot\tan\,(y/x_0)\cdot\frac{\cos^2(y/x_0)}{\cos^2(y/x_0) + \sinh^2\{(x-\epsilon)/x_0\}},\\ \mathscr{R}e\,(V^2) &= \alpha\cdot x + \beta\cdot\tanh\{(x-\epsilon)/x_0\}\cdot\frac{\cosh^2\{(x-\epsilon)/x_0\}}{\cosh^2\{(x-\epsilon)/x_0\} - \sin^2(y/x_0)}. \end{aligned} \tag{39}$$

If δ_c is small we therefore have approximately

$$z_2 = j\cdot y_2 + x_{a_0} = x_{a_0} - j\,\frac{\delta_c}{\gamma_{a_0}}\cdot\frac{1}{1 + \dfrac{\beta}{\gamma_{a_0}x_0}\cdot\dfrac{1}{1+\sinh^2\{(x_{a_0}-\epsilon)/x_0\}}},$$

which means that

$$(40) \qquad |\,T_{0z}\,|^2 \cong \{|\,T_{0z}(\beta = 0)\,|^2\}\,\frac{1}{1 + \dfrac{\beta}{\gamma_{a_0} x_0}\cdot\dfrac{1}{1 + \sinh^2\{(x_{a_0} - \epsilon)/x_0\}}}.$$

We see that only when $(x_{a_0} - \epsilon)/x_0$ is small does the step function increase the transmission coefficient. Even so the resultant effect is significant only if x_0 is small and at the same time $(\epsilon/x_0)^2 \ll 1$. This means in practice that a step in the electron distribution function, in order to be effective in producing the z-wave, must occur almost at the regular interaction level z_{a_0} and it must be quite steep at the same time.

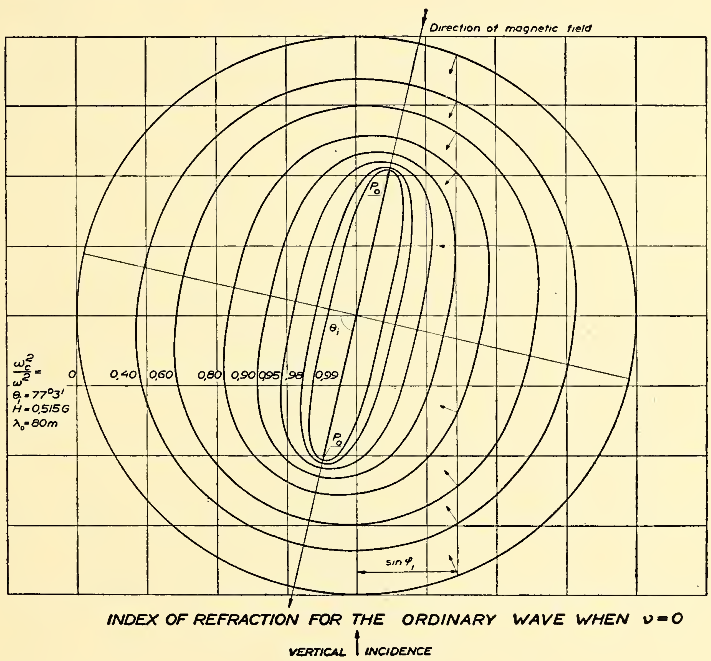


FIGURE 16

Let us next study if such a step, occurring practically at z_{a_0}, is effective in noticeably increasing the transmission coefficient when ν_c is very large, *i.e.* at low geomagnetic latitudes. Because ν_c is very large, $\nu \ll \nu_c$ even for the lower ionosphere. We then have

$$k_0 \int_{z_{a_0}}^{z_2} (n_2 - n_1)\, dz = j \cdot \frac{k_0 \delta_c}{\gamma_{a_0}} \cdot \int_{z_{a_0}}^{z_2} \frac{(n_2 - n_1)\, d\mu}{1 + \dfrac{\beta}{\gamma_{a_0} x_0}\cdot\dfrac{1}{\cosh^2(z/x_0)}},$$

assuming ϵ very small. For a "strong" step function $\beta/x_0 \gg \gamma_{a_0}$ and $\mu \approx (\beta/\delta_c)\tan(z/x_0)$. If $\delta_c \gg \beta$ we find approximately

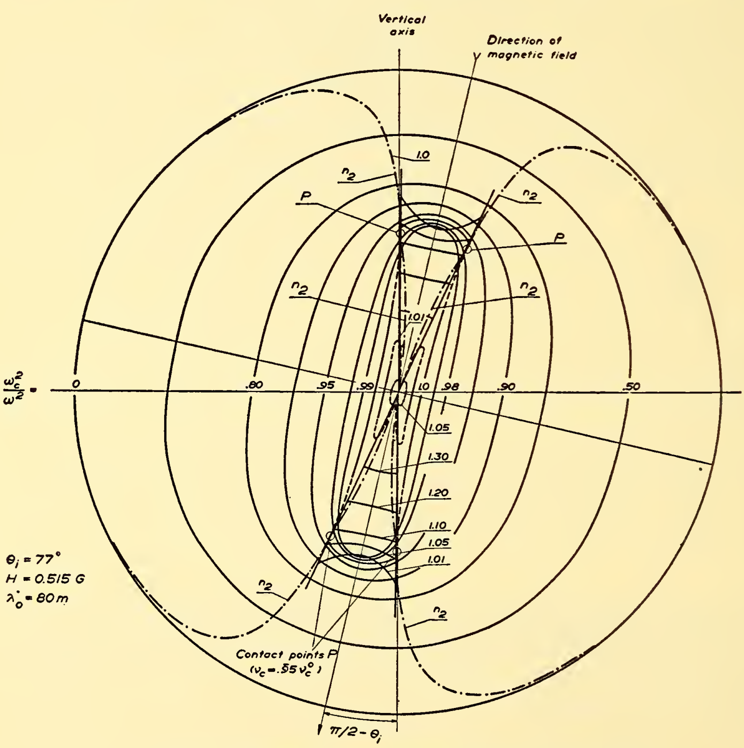
$$
(41)\qquad \begin{aligned} |T_{0z}|^2 &\cong \exp\{-2k_0x_0[\tan^{-1}(\delta_c/\beta) - \tan^{-1}(\delta/\beta)]\} \\ &\cong \exp\{-k_0x_0[\pi - 2\delta/\beta]\} \cong \exp\{-k_0x_0\pi\}, \qquad \delta_c \gg \beta \end{aligned}
$$

We have thus found that "strong" steps in the electron density distribution, located at the interaction level, can produce an interaction or coupling independent of large values of δ_c. Unless the step is located at or very near the interaction level this effect will not occur.

Let us study the situation briefly numerically. If we assume a geomagnetic latitude of about 30° and a wave frequency of $\pi\ Mc/s$, $\delta_c \approx 0.1$. With $\gamma_{a_0} \approx 0.1/km$ (and neglecting μ) our basic formula for the smooth ionosphere yields

$$
|T_{0z}|^2 \cong \exp\{-2k_0\delta_c N(0)/\gamma_{a_0}\} \cong e^{-k_0} \cong e^{-67} \qquad \text{if} \qquad N(0) = 0.5.
$$

In this case a "discontinuity" or step would, therefore, theoretically produce a better z-trace (better $|T_{0z}|$) only if $x_0 < 1/\pi\ km \approx 300\ m \approx 3\lambda_0$. As the trans-



INDEX OF REFRACTION FOR THE ORDINARY WAVE WHEN $\nu = .95\nu_c^0$

FIGURE 17

mission coefficient of the smooth ionosphere is almost infinitely small in the case discussed, one would have to require that $x_0 \approx 30\ m$ in order to produce a

significant result. A "discontinuity" or step coupling therefore yields practical results only if the step takes place within about one vacuum wave length. This requirement, together with the condition that the step be located at the interaction level would make the frequency of "step-excited" triple splits extremely small. It may, however, serve as an explanation of the extremely rare cases reported from more southern latitudes, provided the inospheric irregularities are sufficient.

Finally, one is not surprised to find that exp $\{-k_0 x_0 \pi\}$ also is the amplitude

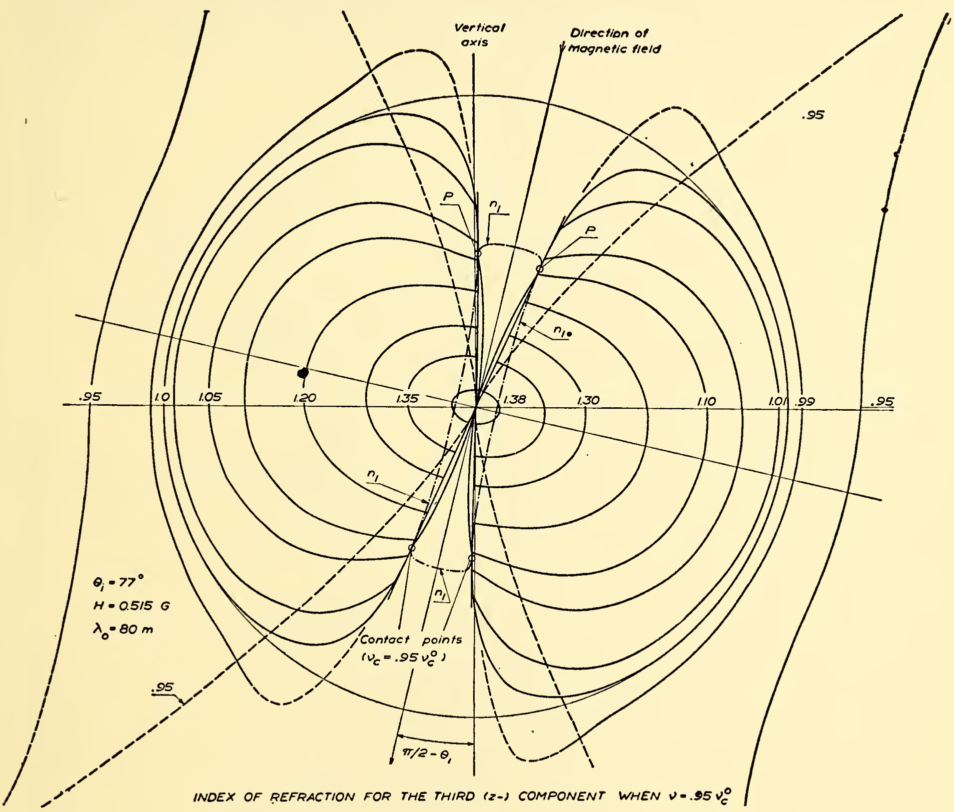


FIGURE 18

of the reflection coefficient of a medium where $\epsilon = 1 + \beta \cdot \tanh(z/x_0)$, [11, p.21], if $\beta \ll 1$, and the wave length is sufficiently short.

6. *Approximate Construction of the Ray Paths*

Many years ago Booker [13] gave a theoretical method of determining the ray paths of the ordinary and extraordinary rays. Here we find it convenient

to use the graphical methods of crystal optics which have recently been used by Poeverlein [14] to determine ray paths in the case of zero losses ($\nu = 0$).

As n is a function of V^2 and φ, the angle that the wave normal makes with the vertical, the geometrical optics requires

$$n(V^2, \varphi) \cdot \sin \varphi = \sin \varphi_i \,. \tag{42}$$

We have thus to find the intersections between the n-curves, plotted as functions of ω_c^2/ω^2 and φ, and the line $\sin \varphi_i$ as shown in Figure 16 for Kiruna conditions for the ordinary ray when $\nu = 0$. The ray or energy direction is normal to the n-curves as shown by the small arrows. It is easily seen that the ray path is

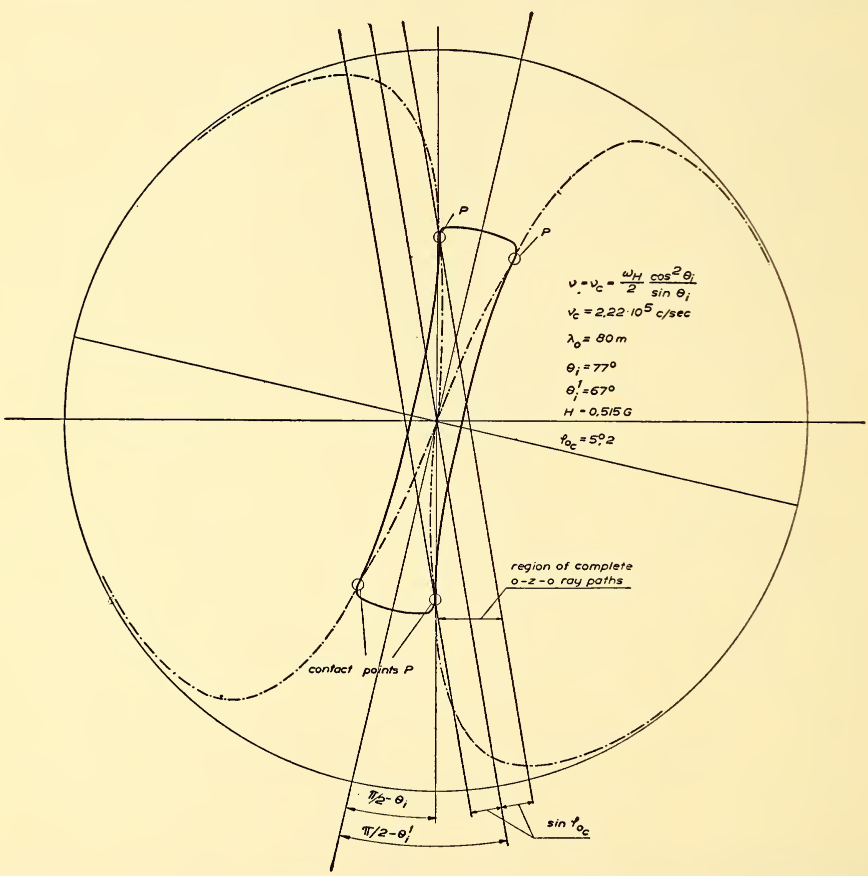


FIG. 19. Depicting the shape of the $n(0, \phi)$- contact curves.

no longer symmetric and the vertically incident ordinary ray is deviated towards the north. The magnitude of this deviation has been the object of a study by S. Forsgren of our laboratory [15].

In order to study the shape of the z-ray we have to plot the $n(V^2, \varphi)$-curves

for $\nu > 0$. The ordinary and z (extraordinary) "ellipses" now become very complicated on account of the fact that for $\nu > \nu_c$ the o-wave transforms to a z-wave (longitudinal transmission). This is clearly shown by Figures 17 and 18 for $n(V^2, \varphi)_0$ and $n(V^2, \varphi)_z$. At P the curves have contact points for $V^2 = 0$. These are the points of practically complete or critical coupling between the o- and z-waves. Here not only the wave normals but also the directions of energy flow are the same.

The shape of the contact curves, $n(0, \varphi)$, is depicted separately in Figure 18 for the sake of convenience. If the inclination θ_i is such that for vertical incidence $\nu_c = \nu$, our vertical line, for the construction of the vertical incidence ray paths, will pass through the two P-points above each other indicating critical

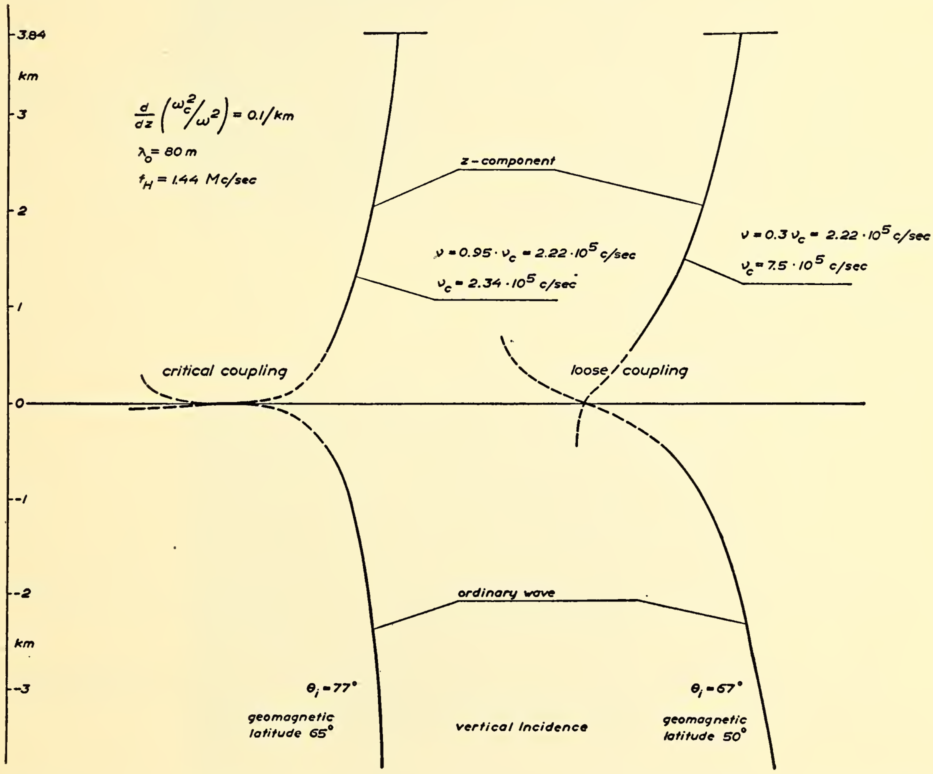


FIG. 20. Ray paths of the o- and z-components when $\nu = 2.22 \cdot 10^5$c/s for geomagnetic latitudes 65° (critical coupling) and 50° (loose coupling).

coupling and practically complete energy transfer for the up-going and for the down-coming wave from z to o. Now even in the case of moderate losses, as in the present cases, the ray paths are quite accurately given by the previous geometrical methods [16], except near the strong coupling regions and the levels of total reflection where the geometrical methods never are applicable anyway.

It is therefore quite easy, making use of Figures 17 and 18, to construct the o- and z-rays for vertical incidence. The result is shown in Figure 20 for vertical incidence and $\nu = 2.22 \cdot 10^5$ c/s at the geomagnetic latitudes 65° and 50°. The characteristic difference between critical and loose coupling is well demonstrated.

One further finds, from Figure 19, that for angles of incidence less than φ_{0c}, one obtains complete o-z-o-rays. Examples of the limiting paths ($\varphi_i = 0$, $\varphi_i = \varphi_{0c}$ north) are shown in Figure 21 for the geomagnetic latitude 50° and, as before, $\nu = 2.22 \cdot 10^5$ c/s. The characteristic distortion of the paths is clearly shown.

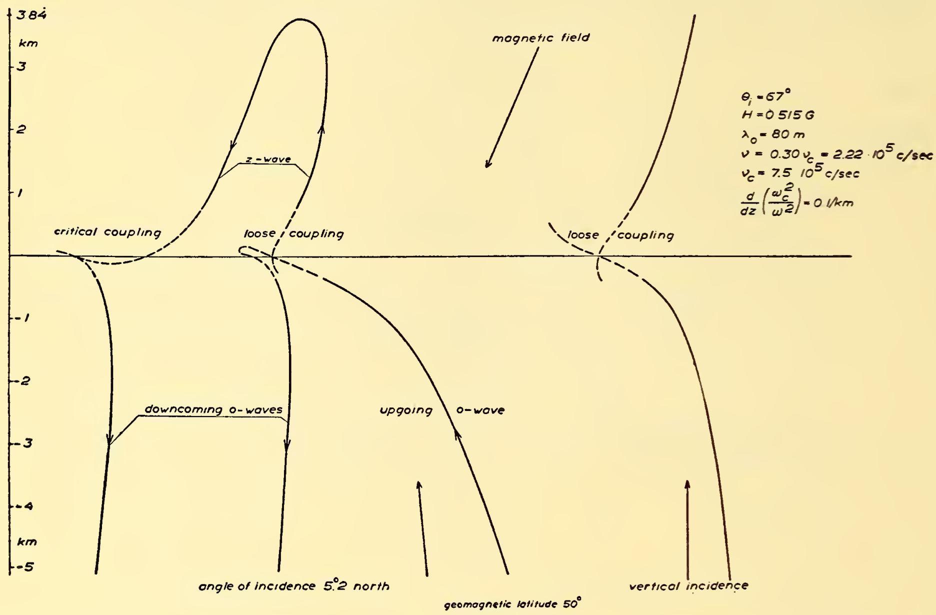


FIG. 21. Ray paths at the limiting oblique incidence, φ_{0c} for $\nu = 2.22 \cdot 10^5$ c/s and a geomagnetic latitude 50°.

The limiting angle, φ_{0c}, normally is quite small. If θ_i be the inclination that would yield $\nu_c = \nu$ and θ_{i_1} the actual inclination, one has

$$\sin \varphi_{0c} = \mathcal{Re}\,(n_p) \cdot \sin\,(\theta_{i_1} - \theta_i), \qquad \nu > \nu_c \tag{43}$$

where

$$\mathcal{Re}\,(n_p) = \{g[g + (1 + g^2)^{1/2}]/2(1 + g^2)\}^{1/2},$$

and

$$g = y \cdot \cos \theta_i\,.$$

For $\theta_{i_1} = 67°$ (geomagnetic latitude 50°) and $\nu = 2.22 \cdot 10^5$ c/s, φ_{0c} is small or only 5.2°.

7. *Experimental Results*

One finds from Figure 14 that, at Kiruna for example, where the geomagnetic latitude is about 65°, triple splitting cannot occur below about 120 km but that it must occur with considerable strength above that level. This is in

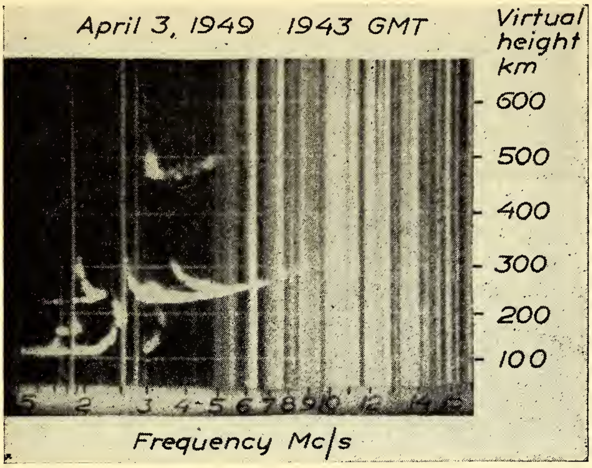


FIG. 22. *E*-layer penetration triple splitting recorded at the Kiruna Observatory, April 3, 1949, 1543 GMT.

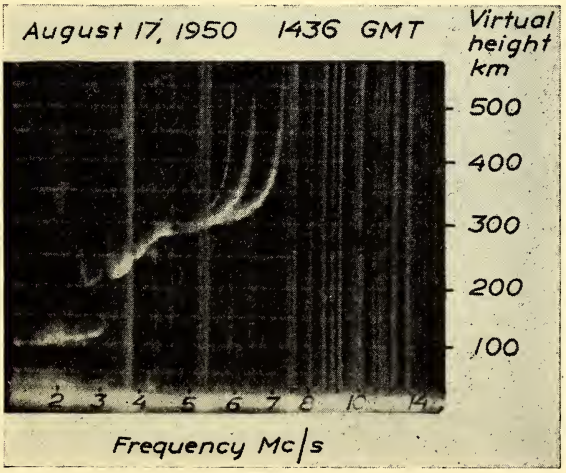


FIG. 23. *E*-layer penetration triple splitting recorded at the Kiruna Observatory, August 17, 1950. (Note the *F* triple split also recorded.)

complete agreement with our experience at this northern observatory where triple penetration of the *E*-layer is recorded almost regularly. In Figure 22 is shown one of the typical records obtained with the 16 kW panoramic recorder

(running without the anti-jamming circuit at the time in question). It is interesting to note that, on this occasion, as on so many others, the *o-z-o*-component is about as strong as the 0-component.

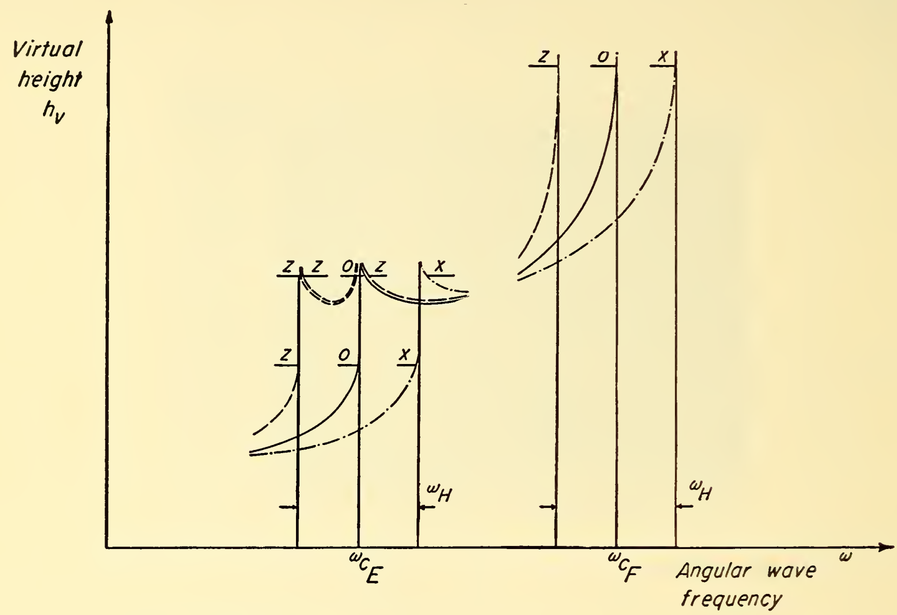


FIG. 24. *z*-, *o*- and *x*-traces for smooth *E*- and *F*-layers.

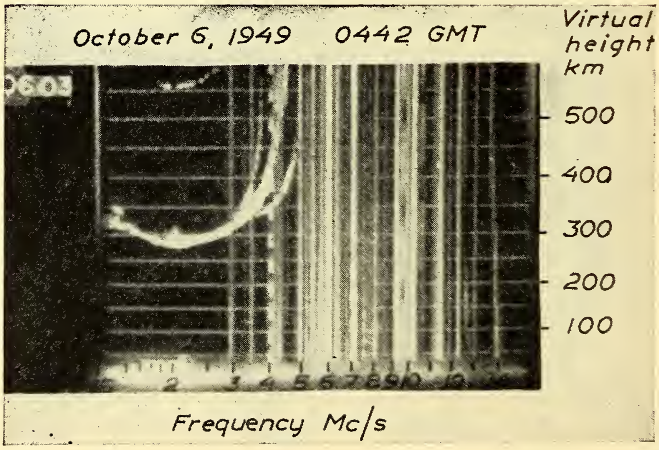


FIG. 25. *F*-layer triple splitting recorded at Kiruna October 6, 1949, 0442 GMT.

If the recorder were moved to an even higher geomagnetic latitude the *o-z-o*-component would be the strongest one. Finally, one notes from Figure 22

the characteristic and obvious fact that the *o-z-o*-wave, reflected from the *F*-layer, disappears when the *o*-wave penetrates the *E*-layer. This feature of the *z*-trace has already been reported by Meek [4]. On certain occasions, however, when the conditions are favourable, the sensitive recording equipment will actually show that this *z*-trace is delayed when the *o*-wave penetrates. Figure 23 is a recent example of this delay obtained at the Kiruna Observatory.

Figure 24, a graphical sketch of the virtual heights for a smooth *E*- and *F*-layer, serves as an additional illustration of the situation.

In Figure 25 we have reproduced a typical average record of a triply split *F*-echo from Kiruna. The *F*-triple splits are somewhat less frequent than the *E*-triple splits at Kiruna as already mentioned.

8. *Concluding Remarks*

The present investigation has shown that with respect to the 0- and *z*-waves the layers of the ionosphere act as if they constituted coupled transmission lines. The "ordinary" lines run from $z_0 - \Delta h$ to z_{a_0} and from $2z_0 - z_{a_0}$ to $z_0 + \Delta h$ for a symmetrical layer. The "*z*" line, adjoining the two "ordinary" lines, runs from z_{a_0} to $2z_0 - z_{a_0}$ ($z_{a_0} < z_0$). When $z_{a_0} = z_0$, *i.e.* for wave frequencies above the *z*-critical frequency, the layer acts as one "ordinary" line and the *z*-trace disappears.

Obviously it is not possible to detect any difference in polarization between the *o*- and *o-z-o*-waves outside the layers. There they are, in fact, both ordinary waves.

The author's thanks are due to the Swedish Research Councils for the Technical and Natural Sciences for grants which made this investigation possible. The excellent assistance is acknowledged of S. Forsgren, E. E., who prepared most of the graphs of this communication.

Bibliography

1. Eckersley, T. L., *Discussion of the ionosphere*, Proceedings of the Royal Society, Series A, Volume 141, 1933.
2. Toshniwal, G. R., *Three-fold magneto ionic splitting of the radio echoes reflected from the ionosphere*, Nature, Volume 135, 1935.
3. Harang, L., *Vertical movements of air in the upper atmosphere*, Terrestrial Magnetism and Atmospheric Electricity, Volume 41, 1936.
4. Meek, J. H., *Triple splitting of ionospheric rays*, Nature, Volume 161, 1948.
5. Newstead, G., *Triple magneto ionic splitting of rays reflected from the F_2 region*, Nature, Volume 161, 1948.
6. Seaton, S. L., *Magneto ionic refraction at high latitudes*, Institute of Radio Engineers Proceedings, Volume 36, 1948.
7. Rydbeck, O. E. H., *The ionospheric and radio wave propagation observatory at Kiruna, 67° 50′ N, 20° 14. 5′E*, Tellus, Volume I, No. 4, 1949.
8. Rydbeck, O. E. H., *Magneto ionic triple splitting of ionospheric waves*, Proceedings of the Indian Academy of Sciences, Volume 31, No. 2, 1950.

9. Rydbeck, O. E. H., *On the propagation of radio waves*, Transactions of the Chalmers University of Technology, Gothenburg, Volume 34, 1944.
10. Appleton, E. V., and Builder, G., *The ionosphere as a doubly refracting medium*, Proceedings of the Physical Society, Volume 45, 1933.
11. Rydbeck, O. E. H., *On the propagation of waves in an inhomogeneous medium*, Report No. 7, Research Laboratory of Electronics, Transactions of the Chalmers University of Technology, Volume 74, 1948.
12. Eckersley, T. L., *Coupling of the ordinary and extraordinary rays in the ionosphere*, Proceedings of the Physical Society, Volume 63, No. 361B, 1950.
13. Booker, H. G., *Oblique propagation of electromagnetic waves in a slowly-varying non-isotropic medium*, Proceedings of the Royal Society, Series A, Volume 155, 1936.
14. Poeverlein, H., *Strahlwege von Radiowellen in der Ionosphäre*, Sitzungsberichte der Bayerischen Akademie der Wissenschaften, 1948.
15. Forsgren, S., *Some calculations of ray paths in the ionosphere*, Report from the Research Laboratory of Electronics 1950, to appear in the Transactions of the Chalmers University of Technology.
16. Rydbeck, O. E. H., *A theoretical survey of the possibilities of determining the distribution of the free electrons in the upper atmosphere*, Transactions of the Chalmers University of Technology, Volume 3, 1942.

An Asymptotic Solution of Maxwell's Equations*

By MORRIS KLINE
New York University

1. *Introduction*

In view of the difficulty of obtaining exact solutions of Maxwell's equations under given initial and boundary conditions and the difficulty of obtaining practically useful solutions even where exact, explicit solutions are known the potentialities of asymptotic solutions warrant investigation. This paper derives a form of asymptotic expansion suited to initial and boundary conditions shortly to be specified and then shows how it is at least theoretically possible to determine the successive coefficients of the expansion through the solution of ordinary differential equations.[1]

A somewhat more specific discussion of the material of this paper follows. Through a form of Duhamel's principle we can relate the electromagnetic field due to an arbitrary electric charge distribution with harmonic time behavior to the field created by the same charge suddenly placed in space at time $t = 0$. The latter field, denoted by $\mathbf{E}_0$, $\mathbf{H}_0$, is to be called the *pulse solution* of Maxwell's equations. Both fields are required to satisfy the initial condition of being zero for $t < 0$ and both are required to satisfy Maxwell's equations for the same electromagnetic parameters ϵ, μ, and σ, the latter being assumed to be sectionally continuous functions of x, y, and z, thereby allowing for abrupt

---

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

---

[1]The form of the asymptotic expansion given in this paper and its derivation from Duhamel's principle were presented by R. K. Luneberg in a series of lectures given at New York University during the academic year 1947–1948. Luneberg derived Duhamel's principle for Maxwell's equations by Laplace transformation of the equations. In this paper the principle is verified directly. Those portions of this paper which repeat Dr. Luneberg's work are presented here for completeness because the material was not published by Dr. Luneberg before his death in 1949. Work on the new material of this paper, which concerns the derivation of the ordinary differential equations for the coefficients of the expansion from discontinuity conditions, was begun by Dr. Luneberg but left incomplete by him.

*The work on this paper was done at the Washington Square College of Arts and Sciences of New York University and was partially supported by Contract No. AF-19(122)-42 with the U. S. Air Force through sponsorship of the Geophysical Research Directorate, Air Force Cambridge Research Laboratories, Air Materiel Command.

changes in media. It is further assumed that any discontinuities with respect to the time variable in $\mathbf{E}_0$, $\mathbf{H}_0$, and their successive time derivatives are finite, an assumption fulfilled in numerous physical situations.

From Duhamel's principle it is possible to show first that the field created by the time harmonic source approaches with increasing time a field having the harmonic time behavior; that is, except for what may be called a transient, the field due to the harmonic source approaches the form $\mathbf{u} \exp \{-i\omega t\}$, $\mathbf{v} \exp \{-i\omega t\}$, where $\mathbf{u}$ and $\mathbf{v}$ are vector functions of x, y, z, and ω is the circular frequency of the harmonic source.

The asymptotic expansions which are the subject of this paper furnish expressions for $\mathbf{u}$ and $\mathbf{v}$ in the form of series in both of which the basic variable is the reciprocal of ω, or, alternatively, the wave length λ. The n-th coefficient of the series for $\mathbf{u}$ is essentially the sum of the discontinuities with respect to the time variable of the $(n - 1)$-st time derivative of $\mathbf{E}_0$. A similar statement applies to $\mathbf{v}$ and $\mathbf{H}_0$. These expansions for $\mathbf{u}$ and $\mathbf{v}$ are noteworthy partly because for the limiting case of λ approaching zero they give the geometrical optics approximation to the time harmonic field.

Since the asymptotic expressions for $\mathbf{u}$ and $\mathbf{v}$ can be obtained from certain discontinuities in the pulse solution one might logically seek that solution. In general, however, the problem of obtaining the pulse solution corresponding to the initial conditions and the conditions on ϵ, μ, and σ is as difficult as solving for the field due to the time harmonic source. But the asymptotic expansions depend only upon certain discontinuities of the pulse solution and this fact suggests the problem of determining these discontinuities directly without requiring a knowledge of the full pulse solution.

To determine the discontinuities directly we first recast Maxwell's differential equations into the form of integral equations which admit a class of discontinuous solutions. Conditions on the discontinuities of these solutions are then obtained which take the form of a recursive system comprising both linear algebraic and linear partial differential equations. This system is recast into a system of recursive ordinary differential equations for the discontinuities of the successive time derivatives of any admissible solution, $\mathbf{E}$, $\mathbf{H}$ of the integral equations. Further conditions in the form of initial conditions on the solutions of the ordinary differential equations are introduced to insure that these solutions give the discontinuities of the pulse field rather than those belonging to some other field.

The entire theory offers a method of obtaining asymptotic solutions for some classes of electromagnetic problems.

2. *Mathematical Statement of the Problem*

We consider, to begin with, the problem of solving Maxwell's equations in a general non-homogeneous isotropic medium, which, for simplicity, is assumed to have zero conductivity. We shall write Maxwell's equations in the form

$$\text{curl}\,\mathbf{H} - \frac{\epsilon}{c}\,\mathbf{E}_t = \frac{1}{c}\,\mathbf{F}_t \tag{1a}$$

$$\text{curl}\,\mathbf{E} + \frac{\mu}{c}\,\mathbf{H}_t = 0, \tag{1b}$$

wherein ϵ and μ are sectionally continuous functions of x, y and z. The real part of $(1/4\pi)\mathbf{F}_t$ represents the enforced current density due to some distribution of charge whose strength varies with time and is to be distinguished from the usual current term σE which represents induced current due to free electrons in conducting media. The function $\mathbf{F}(x, y, z, t)$ shall be considered as given and determines, as we shall see, the charge distribution which creates the field $\mathbf{E}$ and $\mathbf{H}$. The components of $\mathbf{F}$ are assumed to be sectionally smooth functions of x, y, z, t. It is understood that $\mathbf{F} = 0$ for $t < 0$ so that the source begins to act at $t = 0$.

The physical meaning of $\mathbf{F}$ and the justification for including it in equations (1) may be seen by taking the divergence of both sides of the first equation. Since div/curl = 0, we have div $\epsilon\mathbf{E}_t = -\text{div}\ \mathbf{F}_t$. If we assume for the moment[2] that integration with respect to t introduces no new function of x, y, z then div $\epsilon\mathbf{E} = -$ div $\mathbf{F}$. However, by one of the well known electromagnetic equations, div $\mathbf{D} = 4\pi\rho$ where ρ is charge density. Hence $4\pi\rho = -\text{div}\ \mathbf{F}$ and it is clear that $\mathbf{F}$ determines the charge density. From this last equation and the divergence theorem it follows that $\int_V 4\pi\rho\, dv = -\int_A \mathbf{F}\cdot\mathbf{n}\, da$. This relation justifies describing $\mathbf{F}$ as the flux of source charge.

From $4\pi\rho_t = -\text{div}\ \mathbf{F}_t$ and the equation of continuity, $\nabla\cdot\mathbf{J} = -\rho_t$, we have $\nabla\cdot\mathbf{J} = (1/4\pi)\nabla\cdot\mathbf{F}_t$, which means that $(1/4\pi)\mathbf{F}_t = \mathbf{J} =$ current density, except possibly for some divergenceless function of x, y, z. Such a function would have to represent a divergenceless current density which on physical grounds can be assumed not to exist at x, y, z where the source current density $\mathbf{F}_t$ exists.[3]

We shall generally write $\mathbf{F}$ as $\mathbf{g}(x, y, z)f(t)$ where $\mathbf{g}$ is a vector function and f, a scalar, both possibly complex. Such a form for $\mathbf{F}$ is physically reasonable since it means merely that the charge remains at a fixed location but varies in strength with time. Also if $f(t)$ should be the Heaviside unit function $\eta(t)$, that is, 0 for $t < 0$ and 1 for $t > 0$, then the current source in Maxwell's equations is a pulse of current which has infinite strength at $t = 0$ but is nonexistent for other values of t. The field which arises from such a current pulse will be called the *pulse solution* of Maxwell's equations and is denoted by E_0, H_0. Such a field will of course spread out into space from the region in which the source charge exists and will approach with increasing time the static field

---

[2]See the proof in section (4) after equations (21) and (22).

[3]If the term $\sigma\mathbf{E}$ had appeared on the right side of the first of Maxwell's equations, all of the above arguments as to the meaning of $\mathbf{F}$ would apply provided merely that $\mathbf{F}$ exists at points in space at which $\sigma = 0$. This is a reasonable assumption since sources are generally placed in non-conducting media.

created by a fixed charge having the same strength and geometrical distribution. In particular $\mathbf{H}_0(x, y, z, \infty)$ will be zero. Since $\mathbf{F} = 0$ for $t < 0$ it is also true that $\mathbf{E}_0(x, y, z, 0-) = \mathbf{H}_0(x, y, z, 0-) = 0$.

For a time harmonic source we shall let $f(t)$ be $\exp\{-i\omega t\}$ for $t \geq 0$ and 0 for $t < 0$. Here too the resulting field will be zero for $t < 0$.

The function $\mathbf{g}$ which determines the spatial distribution of charge may be accommodated to various physical sources. For a Hertzian dipole $\mathbf{g} = \mathbf{M}\delta(x, y, z)$ where $\mathbf{M}$ is the constant vector moment of the dipole and $\delta(x, y, z)$ is the Dirac δ-function which is infinite at x, y, z and zero elsewhere.

One can now determine a solution of Maxwell's equations for a given $\mathbf{F}$ under the given initial conditions that $\mathbf{E}(x, y, z, t) = \mathbf{H}(x, y, z, t) = 0$ for $t < 0$. However, this solution need not be unique partly because discontinuities in $\mathbf{F}$ are permitted. Discontinuities in $\mathbf{F}$ mean discontinuities in $\mathbf{E}$ and $\mathbf{H}$ at $t = 0+$, for $\mathbf{E}$ and $\mathbf{H}$ are zero at $t = 0+$ where the $\mathbf{g}(x, y, z)$ in $\mathbf{F} = \mathbf{g}f$ is zero, and $\mathbf{E}$ and $\mathbf{H}$ are not zero at $t = 0+$ where $\mathbf{g} \neq 0$. Under discontinuous initial conditions on $\mathbf{E}$ and $\mathbf{H}$ it is possible to show by examples that the resulting solutions of Maxwell's equations need not be unique. In addition, discontinuities in ϵ and μ, which mean a change in medium, call for the imposition of conditions which relate solutions of Maxwell's equations in the differing media. We shall therefore shortly obtain conditions on the discontinuities in $\mathbf{E}$ and $\mathbf{H}$ which are to be satisfied wherever and whenever such discontinuities occur. Under these added conditions the solution is presumably unique.[4]

The problem to which we address ourselves is to find solutions in asymptotic form for Maxwell's equations. These solutions will be fields created by time harmonic sources, are to satisfy the initial condition that $\mathbf{E}$ and $\mathbf{H}$ are zero for $t < 0$, and must correspond to the given, sectionally continuous ϵ and μ. Since we shall show shortly that apart from a transient quantity, that is, a quantity which approaches 0 at each x, y, z with increasing time, the fields are also time harmonic, and indeed possess the same circular frequency as the source, $\mathbf{E}$ and $\mathbf{H}$ will have the form $\mathbf{u}\exp\{-i\omega t\}$ and $\mathbf{v}\exp\{-i\omega t\}$ where $\mathbf{u}$ and $\mathbf{v}$ are complex vector functions of x, y, z. We shall therefore seek asymptotic expressions for $\mathbf{u}$ and $\mathbf{v}$, which give the spatial behavior of the time harmonic part of $\mathbf{E}$ and $\mathbf{H}$. When these asymptotic expressions are obtained we shall offer a method of determining the coefficients of these asymptotic expressions.

3. *Derivation of the Asymptotic Expressions from Duhamel's Principle*

Let $\mathbf{E}_0$, $\mathbf{H}_0$ be the pulse solution of Maxwell's equations for given ϵ and μ and spatial charge distribution $\mathbf{g}$. Let $\mathbf{E}$, $\mathbf{H}$ be the solution corresponding to the same ϵ, μ, and $\mathbf{g}$ but for arbitrary time behavior $f(t)$ of the charge. Both

[4]See the discussion of equations (12) in section 4.

solutions are to satisfy the same initial conditions, that is, 0 for $t < 0$. Then the form of Duhamel's principle which we shall employ states that

$$\mathbf{E} = \frac{\partial}{\partial t} \int_0^t \mathbf{E}_0(t - \tau) f(\tau)\, d\tau \tag{2a}$$

$$\mathbf{H} = \frac{\partial}{\partial t} \int_0^t \mathbf{H}_0(t - \tau) f(\tau)\, d\tau. \tag{2b}$$

While a deductive proof starting from Maxwell's equations and leading to Duhamel's principle can be given it is sufficient for the purposes of this paper to verify the correctness of the principle by direct substitution of **E** and **H** in Maxwell's equations.

One point about this verification needs special attention. On physical grounds it is clear that $\mathbf{E}_0$ and $\mathbf{H}_0$ are certainly discontinuous in the time variable, for, suppose that at $t = 0$ a charge located at some position is suddenly allowed to act. At this time and at any point (x, y, z) distant from the charge the field is zero. However, at some later time, t_1 say, the field created by the charge will reach (x, y, z) and a sudden change will take place there in the value of $\mathbf{E}_0$ and $\mathbf{H}_0$ at time t_1. Thus $\mathbf{E}_0$, $\mathbf{H}_0$, and their successive time derivatives will be discontinuous at $t = t_1$. If ϵ and μ are discontinuous along some surface, then the field upon reaching that surface will be partially reflected and the reflected field may pass (x, y, z) at some later time, t_2 say, at which time another discontinuity will exist in the values of $\mathbf{E}_0$, and $\mathbf{H}_0$, and their successive time derivatives. In addition, discontinuities in $\mathbf{E}_0$ and $\mathbf{H}_0$ will exist at time $t = 0$ at those (x, y, z) at which charges are placed, for at $t = 0-$, the field is 0 but at $t = 0+$ the field created by the charge arises immediately.

In verifying Duhamel's principle it is therefore necessary to take into account discontinuities in $\mathbf{E}_0$ and $\mathbf{H}_0$. Let (x_1, y_1, z_1, t_1) be a point at which $\mathbf{E}_0$ and $\mathbf{H}_0$ are discontinuous and let t_1 be such that $0 < t_1 < t$. The verification is accomplished by showing that the indefinite time integrals of **E** and **H** satisfy the time integrals of Maxwell's equation. We consider the first of Maxwell's equations and write

$$\operatorname{curl} \int \mathbf{H}\, dt - \frac{\epsilon}{c} \mathbf{E} = \frac{1}{c} \mathbf{F}.$$

We may now substitute (2a) and (2b) in this equation to obtain

$$\operatorname{curl} \int_0^t \mathbf{H}_0(t - \tau) f(\tau)\, d\tau - \frac{\epsilon}{c} \frac{\partial}{\partial t} \int_0^t \mathbf{E}_0(t - \tau) f(\tau)\, d\tau = \frac{1}{c} \mathbf{F}. \tag{3}$$

Consider the second integral. Since $\mathbf{E}_0(t)$ is discontinuous at $t = t_1$, $\mathbf{E}_0(t - \tau)$ is discontinuous for $\tau = t - t_1$. Let us break up the τ-interval of integration, 0 to t, into 0 to $t - t_1$ and $t - t_1$ to t. Within these subintervals $\mathbf{E}_0(t - \tau)$ is a

continuous function of τ. Moreover we may suppose on physical grounds that $\mathbf{E}_0$ and $\partial \mathbf{E}_0/\partial t$ are continuous at $t = t$. Hence[5] we may differentiate the separate integrals with respect to t and obtain

$$-\frac{\epsilon}{c}\int_0^t \mathbf{E}_{0t}(t - \tau)f(\tau)\,d\tau - \frac{\epsilon}{c}\mathbf{E}_0(t_1^+)f(t - t_1) - \frac{\epsilon}{c}\mathbf{E}_0(0)f(t) + \frac{\epsilon}{c}\mathbf{E}_0(t_1^-)f(t - t_1).$$

We consider next the first integral in (3). Since the curl operation also involves a set of differentiations under the integral sign we must again consider the discontinuity in $\mathbf{H}_0(t)$, which we may with no loss in generality suppose to occur at $t = t_1$. We must now note that in later applications t_1 will be a function of x, y, z and we may write the equation of the hyper-surface on which the discontinuities x, y, z, t lie in the form $t_1 = \psi(x, y, z)/c$. We again break up the τ-range of the integral into 0 to $t - t_1$ and $t - t_1$ to t. Considering now only one of the differentiations involved in the curl operation and designating by $\mathbf{H}_{03}$ the third component of $\mathbf{H}_0$, we have

$$\frac{\partial}{\partial y}\int_0^t \mathbf{H}_{03}(t - \tau)f(\tau)\,d\tau$$

$$= \int_0^t \frac{\partial}{\partial y}\mathbf{H}_{03}(t - \tau)f(\tau)\,d\tau - H_{03}(t_1^+)f(t - t_1)\frac{\partial t_1}{\partial y} + \mathbf{H}_{03}(t_1^-)f(t - t_1)\frac{\partial t_1}{\partial y}.$$

If we now take into account all of the terms involved in the curl operation we obtain

$$\operatorname{curl}\int_0^t \mathbf{H}_0(t - \tau)f(\tau)\,d\tau$$

$$= \int_0^t \operatorname{curl}\mathbf{H}_0(t - \tau)f(\tau)\,d\tau - \frac{1}{c}\operatorname{grad}\psi \times [\mathbf{H}_0(t_1)]f(t - t_1),$$

wherein $[\mathbf{H}_0(t_1)]$ equals $\mathbf{H}_0(t_1+) - \mathbf{H}_0(t_1-)$.

If we substitute the two results of differentiation in (3) and use the fact that $\mathbf{E}_0$ and $\mathbf{H}_0$ are a solution of Maxwell's equations for the pulse source $\mathbf{F} = \mathbf{g}\eta$ we obtain

$$\int_0^t \frac{1}{c}\mathbf{g}\eta'(t - \tau)f(\tau)\,d\tau$$

$$- \frac{1}{c}\operatorname{grad}\psi \times [\mathbf{H}_0(t_1)]f(t - t_1) - \frac{\epsilon}{c}[\mathbf{E}_0(t_1)]f(t - t_1) - \frac{\epsilon}{c}\mathbf{E}_0(0)f(t) = \frac{1}{c}\mathbf{g}f(t).$$

---

[5]See P. Franklin, *Treatise on Advanced Calculus*, Wiley, New York, 1940, p. 348. We are supposing, too, that the discontinuity in $\mathbf{E}_{0t}(t)$ at $t = t_1$ is finite. This assumption is consistent with later work in this paper.

We shall show later[6] that $\mathbf{E}_0(0+) = -\mathbf{g}/\epsilon$. Since $\eta'(t) = 0$ for $t \geq 0+$, the integral is 0. We therefore arrive at the condition

$$\operatorname{grad} \psi \times [\mathbf{H}_0] + \epsilon[\mathbf{E}_0] = 0$$

on the discontinuity hypersurface $t_1 = \psi/c$ as the condition that the **E** and **H** given by (2a) and (2b) satisfy the first of Maxwell's equations.

A similar discussion to show that this **E** and **H** satisfy the second of Maxwell's equation would lead to the condition

$$\operatorname{grad} \psi \times [\mathbf{E}_0] - \mu[\mathbf{H}_0] = 0$$

on the hypersurface $t_1 = \psi/c$.

We require that $\mathbf{E}_0$ and $\mathbf{H}_0$ satisfy the above conditions on a discontinuity surface and the verification of Duhamel's principle is thereby completed. These conditions are in no sense a limitation on Duhamel's principle since some conditions must be placed on $\mathbf{E}_0$ and $\mathbf{H}_0$ to relate solutions of Maxwell's equations across a discontinuity surface. We shall see later that the above conditions are precisely the ones we must impose.

We may now derive from Duhamel's principle the asymptotic expansion with which this paper deals. Since the expansion applies only to fields created by time harmonic sources we shall now let $f(t) = \exp\{-i\omega t\}$. Let $t - \tau = s$ and $t = t$ in (2a). By this change of variable we obtain

$$\mathbf{E} = \frac{\partial}{\partial t}\int_0^t \mathbf{E}_0(s) \exp\{-i\omega(t - s)\}\, ds.$$

Taking the $\exp\{-i\omega t\}$ factor outside the integral and differentiating the product gives

$$\mathbf{E} = \mathbf{E}_0(t) - i\omega \exp\{-i\omega t\} \int_0^t \mathbf{E}_0(s) \exp\{i\omega s\}\, ds.$$

Adding and subtracting equivalent terms gives

$$\begin{aligned}\mathbf{E} &= \mathbf{E}_0(t) - \mathbf{E}_0(\infty)\\ &\quad - \exp\{-i\omega t\}\left\{-\mathbf{E}_0(\infty) + i\omega \int_0^t (\mathbf{E}_0(s) - \mathbf{E}_0(\infty)) \exp\{i\omega s\}\, ds\right\}.\end{aligned}$$

This result shows that with increasing time $\mathbf{E}(x, y, z, t)$ approaches[7] the form $\mathbf{u} \exp\{-i\omega t\}$ where, changing s to τ in the preceding result,

$$\text{(4a)} \qquad \mathbf{u}(x, y, z) = \mathbf{E}_0(\infty) - i\omega \int_0^\infty (\mathbf{E}_0(\tau) - \mathbf{E}_0(\infty)) \exp\{i\omega\tau\}\, d\tau.$$

---

[6]See footnote 11. We also show later that $\mathbf{H}_0(0+) = 0$. This fact is needed in verifying that **E** and **H** satisfy the second Maxwell equation.

[7]The existence of the infinite integral in (4a) may be assumed on the ground that $\mathbf{E}_0(\tau)$ approaches, with larger and larger τ, the field of a static charge and hence the integrand must approach zero sufficiently rapidly for convergence.

That is, a time harmonic source gives rise, except for a transient, to a time harmonic field of the same frequency.

Similarly, $\mathbf{H}$ approaches the form $\mathbf{v} \exp\{-i\omega t\}$ where

$$\text{(4b)} \qquad \mathbf{v}(x, y, z) = \mathbf{H}_0(\infty) - i\omega \int_0^\infty (\mathbf{H}_0(\tau) - \mathbf{H}_0(\infty)) \exp\{i\omega\tau\}\, d\tau.$$

Now integrate (4a) by parts, letting the u of the parts formula be $\mathbf{E}_0(\tau) - \mathbf{E}_0(\infty)$ and the dv be $i\omega \exp\{i\omega\tau\}\, d\tau$. Then

$$\mathbf{u} = \mathbf{E}_0(\infty) - (\mathbf{E}_0(\tau) - \mathbf{E}_0(\infty)) \exp\{i\omega\tau\} \Big|_0^\infty + \int_0^\infty \mathbf{E}_{0\tau}(\tau) \exp\{i\omega\tau\}\, d\tau.$$

In evaluating the second term on the right side we must take into account the fact that $\mathbf{E}_0(\tau)$ is discontinuous. Let $\tau_1, \tau_2, \cdots, \tau_\alpha, \cdots, \tau_n$ be values of τ at which $\mathbf{E}_0(\tau)$ is discontinuous.[8] Further, let $[\mathbf{E}_0]_\alpha = \mathbf{E}_0(\tau_\alpha +) - \mathbf{E}_0(\tau_\alpha -)$. Then

$$\mathbf{u} = \mathbf{E}_0(0+) + \sum_\alpha [\mathbf{E}_0]_\alpha \exp\{i\omega\tau_\alpha\} + \int_0^\infty \mathbf{E}_{0\tau}(\tau) \exp\{i\omega\tau\}\, d\tau.$$

We again integrate by parts obtaining

$$\mathbf{u} = \mathbf{E}_0(0+) + \sum_\alpha [\mathbf{E}_0] \exp\{i\omega\tau_\alpha\}$$

$$- \frac{1}{i\omega} \sum_\alpha [\mathbf{E}_{0\tau}]_\alpha \exp\{i\omega\tau_\alpha\} - \frac{1}{i\omega} \int_0^\infty \mathbf{E}_{0\tau\tau} \exp\{i\omega\tau\}\, d\tau.$$

The second summation includes strictly a discontinuity at $\tau = 0$ and $\tau = \infty$. However, we shall show[9] later that except at special points $\mathbf{E}_{0\tau}(0+) = 0$; also we may assume that $\mathbf{E}_{0\tau}(\infty) = 0$ on physical grounds since $\mathbf{E}_0(\tau)$ approaches the field of a static charge as τ becomes infinite. The same remarks apply to the higher derivatives of $\mathbf{E}_0(\tau)$. $\mathbf{E}_0(0+)$ has already been noted to equal $-\mathbf{g}/\epsilon$ and so may be replaced by that quantity.

It is apparent that the process of integration by parts can be continued as long as the next higher time derivative of $\mathbf{E}_0$ exists and as long as the discontinuities of the time derivative of $\mathbf{E}_0$ which appears in the integrand are finite. These conditions are certainly fulfilled for large classes of physical problems.

---

[8]$\mathbf{E}_0(\tau)$ may be discontinuous at an infinite number of values of τ. In such cases the summations which occur in the succeeding steps must be taken over the infinite number of values of α. On physical grounds these sums may be expected to converge, for the physical circumstance under which an infinite series of discontinuities occurs is that of a series of reflections of a family of wave fronts running between two refracting surfaces, e.g., internal reflections in a lens. At each reflection the value of $\mathbf{E}_0$ and hence the next discontinuity in $\mathbf{E}_0$ which occurs at (x, y, z) may be expected to decrease sufficiently rapidly to insure convergence.

[9]See footnote 11.

We may therefore write

$$
\begin{aligned}
\mathbf{u} = -\mathbf{g}/\epsilon + \sum_\alpha [\mathbf{E}_0]_\alpha \exp\{i\omega\tau_\alpha\} \\
\text{(5)} \qquad - \frac{1}{i\omega} \sum_\alpha [\mathbf{E}_{0\tau}]_\alpha \exp\{i\omega\tau_\alpha\} + \frac{1}{(i\omega)^2} \sum_\alpha [\mathbf{E}_{0\tau\tau}]_\alpha \exp\{i\omega\tau_\alpha\} - \cdots .
\end{aligned}
$$

The summation which appears in each term must cover all the discontinuities which exist for the derivative in question.

The infinite series in (5) is truly an asymptotic expression for $\mathbf{u}$, for, the remainder after n terms is readily seen, by an integration by parts, to be of order one higher in $1/i\omega$ than the n-th term. On the assumption that the integral which results from the integration by parts remains finite as ω becomes infinite, the definition of an asymptotic expansion is satisfied. In view, however, of the physical meaning of $\mathbf{E}_0(\tau)$ the time derivatives may be expected to approach zero sufficiently rapidly.

The series (5) can be given a slightly altered form. The points (x, y, z, t) at which $\mathbf{E}_0$, $\mathbf{H}_0$, and their successive time derivatives are discontinuous in t may be regarded as lying on some complicated hypersurface $\phi(x, y, z, t) = 0$. We assume on physical grounds that it is possible to solve for t. This solution may have many branches, $t_\alpha = \psi_\alpha(x, y, z)/c$. As a matter of fact, at points where ϵ, μ, and $\mathbf{F}$ are continuous we shall show later that the ψ_α are wave-fronts, that is, they satisfy the eiconal equation $\psi_x^2 + \psi_y^2 + \psi_z^2 = \epsilon\mu$. Now, in (5), x, y, and z are suppressed variables. For any given (x, y, z) at which $\mathbf{u}$ is being evaluated, there are various values t_α at which $\mathbf{E}_0$ or some time derivative is discontinuous. At that x, y, z the t_α are given by $\psi_\alpha(x, y, z)/c$. Hence we may write (5) as

$$
\begin{aligned}
\mathbf{u} = -\mathbf{g}/\epsilon + \sum_\alpha [\mathbf{E}_0]_\alpha \exp\{ik\psi_\alpha\} - \frac{1}{i\omega} \sum_\alpha [\mathbf{E}_{0\tau}]_\alpha \exp\{ik\psi_\alpha\} \\
\text{(6a)} \qquad + \frac{1}{(i\omega)^2} \sum_\alpha [\mathbf{E}_{0\tau\tau}]_\alpha \exp\{ik\psi_\alpha\} - \cdots .
\end{aligned}
$$

By the same argument we obtain the asymptotic expansion for $\mathbf{v}$ except for the fact[10] that $\mathbf{H}_0(0+) = 0$. Hence

$$
\begin{aligned}
\mathbf{v} = \sum_\alpha [\mathbf{H}_0]_\alpha \exp\{ik\psi_\alpha\} - \frac{1}{i\omega} \sum_\alpha [\mathbf{H}_{0\tau}]_\alpha \exp\{ik\psi_\alpha\} \\
\text{(6b)} \qquad + \frac{1}{(i\omega)^2} \sum_\alpha [\mathbf{H}_{0\tau\tau}]_\alpha \exp\{ik\psi_\alpha\} - \cdots .
\end{aligned}
$$

---

[10]See footnote 11.

One point about the expansions for **u** and **v** that is especially noteworthy is the significance of the leading terms. Ignoring the term $-\mathbf{g}/\epsilon$ in (6a) since this term is zero except at points where source charge exists, we note that for $\omega \to \infty$ or $\lambda \to 0$ the first term in each expansion should be a good approximation to the spatial behavior of the time harmonic field. These first terms are the approximation given by geometrical optics.

Formulas (6a) and (6b) constitute a method of finding the asymptotic expansion of a large class of time harmonic fields. In particular they apply to those physical situations where discontinuities in $\mathbf{E}_0$ and $\mathbf{H}_0$ arise from the sudden passage past any point x, y, z of wave fronts created by the source and coming either directly from the source or reflected from other discontinuities in the medium. In these situations the various assumptions made above, and therefore the asymptotic expansion, apply.

One can use formulas (6) if the pulse solution $\mathbf{E}_0$, $\mathbf{H}_0$ is known. It is possible to obtain $\mathbf{E}_0$ and $\mathbf{H}_0$ in some classes of problems, but, generally, finding $\mathbf{E}_0$ and $\mathbf{H}_0$ is as difficult a problem as finding **E** and **H**, or **u** and **v**, directly. However, formulas (6) show that the coefficients of the asymptotic expansion depend directly upon the behavior of $\mathbf{E}_0$ and $\mathbf{H}_0$ only in the immediate neighborhood of the discontinuity surfaces $t_\alpha = \psi_\alpha(x, y, z)/c$. Hence the problem suggests itself to determine these coefficients directly instead of finding the complete $\mathbf{E}_0$ and $\mathbf{H}_0$. The coefficients we seek are essentially discontinuities of a solution of Maxwell's equation and we shall therefore investigate these discontinuities. The procedure we shall adopt is to investigate the discontinuities of a large class of solutions which includes pulse solutions and we shall obtain conditions which discontinuities of **E**, **H**, and the time derivatives of **E** and **H** must satisfy. These discontinuity conditions will then be used to determine the coefficients of the expansions for **u** and **v**.

4. *Derivation of the Discontinuity Conditions*

We shall be interested in fields **E** and **H** which are discontinuous along some hypersurface $\phi = 0$, which satisfy Maxwell's equations on either side of this surface, and which satisfy given initial conditions. Discontinuous solutions arise because the initial conditions are discontinuous and discontinuities in the medium are permitted. Apparently some conditions must relate the two fields which obtain on either side of the discontinuity surface else no unique solution is determined by the initial conditions.

To treat discontinuous solutions of Maxwell's equations we shall replace these equations by integral equations in which **E**, **H**, and **F**, but not their derivatives, will appear. We shall then treat solutions of these integral equations and, in particular, derive discontinuity conditions from them. Further advantages which result from working with the integral equations will be discussed later.

Consider first any region G of x, y, z, t space in which **E**, **H**, and all their derivatives exist. Let $\Omega(x, y, z, t)$ be any scalar function which is continuous and has continuous partial derivatives of any finite order in G and which is identically 0 on the boundary and outside of G. Let us multiply each of the equations (1) by Ω and integrate over G. If for brevity we let

$$\mathbf{C} = \frac{\epsilon}{c}\mathbf{E} + \frac{1}{c}\mathbf{F}, \tag{7}$$

we obtain

$$\int_G (\Omega \operatorname{curl} \mathbf{H} - \Omega \mathbf{C}_t)\, dw = 0, \tag{8a}$$

$$\int_G \left(\Omega \operatorname{curl} \mathbf{E} + \Omega \frac{\mu}{c} \mathbf{H}_t\right) dw = 0, \tag{8b}$$

where by $\int_G$ we mean the four-fold integral and $dw = dx\, dy\, dz\, dt$.

Now, using H_1, H_2, H_3 for the components of **H** and literal subscripts to indicate partial derivatives, we have for the first term in (8a),

$$\int_G \Omega \operatorname{curl} \mathbf{H}\, dw = \int_G [\Omega(H_{3_y} - H_{2_z})i + \text{corresponding terms}]\, dw. \tag{9}$$

We apply integration by parts to each term with respect to the variable appearing in the partial derivative. Let Γ be the boundary of G. Then the integration by parts applied to the first term gives

$$\int_G \Omega H_{3_y}\, dw = \int_\Gamma \Omega H_3\, dx\, dz\, dt - \int_G \Omega_y H_3\, dw.$$

The $\int_\Gamma$ denotes that we must replace the independent variable y in Ω and H_3 by $h(x, z, t)$ and then by $k(x, z, t)$, h and k being the equations of the two parts of Γ which bound G, and then subtract the result of the second substitution from the first. However, since Ω is 0 on the boundary of G the triple integral vanishes. Application of this procedure of integration by parts to each term of (9) yields

$$\int_G \Omega \operatorname{curl} \mathbf{H}\, dw = -\int_G \operatorname{grad} \Omega \times \mathbf{H}\, dw, \tag{10}$$

where the curl gradient operators are the usual three-dimensional operators.

Next consider the second term of (8a). The same procedure of integration by parts, this time with respect to the variable t, yields

$$\int_G \Omega \mathbf{C}_t\, dw = -\int_G \Omega_t \mathbf{C}\, dw. \tag{11}$$

From equation (1a) we have therefore derived the integral equation

$$\int_G (\text{grad } \Omega \times \mathbf{H} - \Omega_t \mathbf{C})\, dw = 0. \tag{12a}$$

Similarly from equation (1b) we obtain

$$\int_G (\text{grad } \Omega \times \mathbf{E} + \Omega_t (\mu/c) \mathbf{H})\, dw = 0. \tag{12b}$$

Now in a region G in which **E**, **H**, their first order derivatives, **F**, and $\mathbf{F}_t$ exist the equations (1) are equivalent to equations (12), for not only have we derived (12) from (1) but we may derive (1) from (12) by proceeding first to (8) and then arguing that since (8) must hold for *any* function Ω of the kind described above, the factor multiplying Ω in the integrand must be 0.

If, however, we consider a region G in which the first order derivatives of **E** and **H** are discontinuous and possibly **E**, **H**, **F**, and $\mathbf{F}_t$ likewise discontinuous, then equations (1) do not hold throughout the region whereas equations (12) do still have significance. Hence we agree to *replace* Maxwell's equation (1) by the integral equations (12) for all future work in this paper, the purpose of this move being to admit discontinuous solutions of Maxwell's equations.

One further condition will be attached to the use of the integral equations (12). Suppose that in the region G, the functions **E**, **H**, **F** or their derivatives are discontinuous at points (x, y, z, t) which lie on some hypersurface $\phi = 0$ contained or partially contained in G (see Fig. 1). We shall suppose that **E**, **H** and **F** satisfy Maxwell's equations, and in fact have higher derivatives with

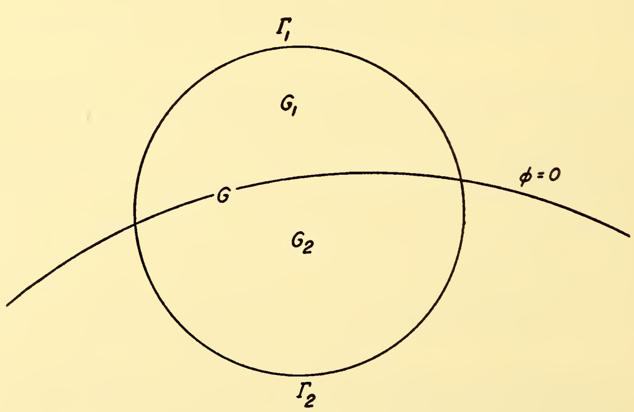


FIGURE 1

respect to t, in and on the boundary of G_1 and G_2 separately, the values of **E**, **H**, **F** and their derivatives on the boundary being the limiting values approached by these functions from the interior.

Imposition of the requirement that solutions of the integral equations (12) valid over the domain G also satisfy Maxwell's differential equations in G_1 and G_2 separately raises the question of whether such solutions exist at all. It can be shown mathematically that if a discontinuous solution $\mathbf{E}$ and $\mathbf{H}$ is the limit of a sequence of continuous solutions $\mathbf{E}_n$, $\mathbf{H}_n$ then this discontinuous solution will satisfy the integral equations (12) in G and Maxwell's equations (1) in G_1 and G_2 separately. That is, there are solutions which meet our conditions and it is for such solutions that we shall obtain discontinuity conditions.

There is good physical reason to expect that discontinuous solutions are limits of continuous solutions. For example, a sudden change in the physical characteristics of a medium, which is idealized mathematically as a discontinuous change in the characteristics, can be replaced by a rapidly varying continuous change to which a continuous solution would correspond. As the continuous change is made more and more rapid so as to approach the sudden change, the corresponding continuous solutions of the field equations should approach the discontinuous solution corresponding to the sudden change. Hence one expects discontinuous solutions of the field equations to be limits of continuous solutions.

There are several advantages to replacing equations (1) by the integral equations (12). First, we have generalized Maxwell's equations so that we may take care of discontinuities in $\mathbf{E}$, $\mathbf{H}$, $\mathbf{F}$, and their derivatives. *Equations* (12) *hold as long as the components of* $\mathbf{E}$, $\mathbf{H}$, *and* $\mathbf{F}$ *are integrable.* Secondly, we have reason (though no rigorous proof) to expect the uniqueness of the solutions to equations (12) under the initial conditions stated in article 2, for the solutions we consider are limits of continuous solutions and the latter are unique for continuous initial conditions.

A third advantage of equations (12), which at the moment is incidental, is that it permits us to treat without difficulty point source functions of the form $\mathbf{F} = \mathbf{M}\ \delta(x, y, z)f(t)$, where $\mathbf{M}$ is a constant vector and δ is the Dirac delta function. For this $\mathbf{F}$, using a well known property of the delta function, equation (12a) reads

$$\int_G [\operatorname{grad} \Omega \times \mathbf{H} - \frac{\epsilon}{c}\, \Omega_t \mathbf{E}]\, dw = \frac{\mathbf{M}}{c} \int_0^\infty \Omega_t(0, 0, 0, t) f(t)\, dt.$$

We shall now apply equations (12) to a domain G within which discontinuities occur. Let $\phi(x, y, z, t) = 0$ be any surface dividing the region G (see Figure 1) into regions G_1 and G_2 and such that $\mathbf{F}$, $\mathbf{E}$, $\mathbf{H}$, or some of their derivatives are discontinuous on this surface. However, in G_1 and G_2 and on the boundaries as approached from G_1 or G_2 let $\mathbf{E}$, $\mathbf{H}$ and $\mathbf{F}$ have derivatives of any order.

We may regard the fields $\mathbf{E}$ and $\mathbf{H}$ which are defined in G to consist of two fields. In G_1 , $\mathbf{E}$ and $\mathbf{H}$ are to be the fields $\mathbf{E}^{(1)}$ and $\mathbf{H}^{(1)}$ which have the same values as $\mathbf{E}$ and $\mathbf{H}$ in G_1 and which are 0 in G_2 . On $\phi = 0$, $\mathbf{E}^{(1)}$ and $\mathbf{H}^{(1)}$ have the values approached by $\mathbf{E}$ and $\mathbf{H}$ from G_1 . Likewise in G_2 , $\mathbf{E}$ and $\mathbf{H}$

are to be replaced by $\mathbf{E}^{(2)}$ and $\mathbf{H}^{(2)}$ which are to have the same values as $\mathbf{E}$ and $\mathbf{H}$ there but which are zero in G_1. Likewise, on $\phi = 0$, $\mathbf{E}^{(2)}$ and $\mathbf{H}^{(2)}$ are to have the values approached by $\mathbf{E}$ and $\mathbf{H}$ from G_2. With this understanding about the values of $\mathbf{E}$ and $\mathbf{H}$ in G_1 and G_2 we may now state instead of (12a), for example,

$$\begin{aligned}&\int_G (\text{grad } \Omega \times \mathbf{H} - \Omega_t \mathbf{C})\, dw \\ (13) \qquad &= \int_{G_1} (\text{grad } \Omega \times \mathbf{H} - \Omega_t \mathbf{C})\, dw + \int_{G_2} (\text{grad } \Omega \times \mathbf{H} - \Omega_t \mathbf{C})\, dw.\end{aligned}$$

The same applies to equation (12b).

Let us now assume that we have a field $\mathbf{E}, \mathbf{H}$ which is a solution of equations (12) but which is discontinuous on the hypersurface $\phi = 0$ separating G into G_1 and G_2. We shall apply equations (12) in the form indicated by the right side of equation (13) to derive conditions on the discontinuities on $\mathbf{E}$ and $\mathbf{H}$.

We first introduce the unit vector $\mathbf{N}$ which is normal to the surface $\phi = 0$. The components of $\mathbf{N}$ are

$$(14) \qquad \mathbf{N} = \pm\lambda(\phi_x, \phi_y, \phi_z, \phi_t)$$

where $\lambda = 1/(\phi_x^2 + \phi_y^2 + \phi_z^2 + \phi_t^2)^{1/2}$, the plus indicating that $\mathbf{N}$ points into G_2 and the minus that $\mathbf{N}$ points into G_1. Now consider, as in the derivation of (10), a typical term of $\int_{G_1} \text{grad } \Omega \times \mathbf{H}\, dw$. Such a term is $\int_{G_1} \Omega_y H_3\, dw$. By integration by parts with respect to y we obtain that

$$\int_{G_1} \Omega_y H_3\, dw = \int_{\Gamma_1 + \phi} \Omega H_3\, dx\, dz\, dt - \int_{G_1} \Omega H_{3_y}\, dw,$$

wherein Γ_1 is that part of Γ which forms part of the boundary of G_1. Since $\Omega = 0$ on Γ,

$$\int_{G_1} \Omega_y H_3\, dw = \int_{\phi} \Omega H_3\, dx\, dz\, dt - \int_{G_1} \Omega H_{3_y}\, dw.$$

Since $dx\, dz\, dt$ is the projection of an element of surface of ϕ on the x, z, t space, it may be replaced by $\lambda\phi_y\, ds$ where ds is an element of ϕ. If we apply the process just indicated to each term of $\int_{G_1} \text{grad } \Omega \times \mathbf{H}\, dw$ we obtain that

$$(15) \qquad \int_{G_1} \text{grad } \Omega \times \mathbf{H}\, dw = \int_{\phi} \Omega \text{ grad } \phi \times \mathbf{H}\lambda\, ds - \int_{G_1} \Omega \text{ curl } \mathbf{H}\, dw.$$

Similarly, for values of $\mathbf{H}$ within and on the boundary of G_2,

$$(16) \qquad \int_{G_2} \text{grad } \Omega \times \mathbf{H}\, dw = \int_{\phi} \Omega \text{ grad } \phi \times \mathbf{H}(-\lambda)\, ds - \int_{G_2} \Omega \text{ curl } \mathbf{H}\, dw.$$

Likewise, using integration by parts with respect to the variable t, we obtain that

$$(17) \qquad -\int_{G_1} \Omega_t \mathbf{C}\, dw = -\int_{\phi} \Omega \mathbf{C} \lambda \phi_t\, ds + \int_{G_1} \Omega \mathbf{C}_t\, dw.$$

Similarly,

$$(18) \qquad -\int_{G_2} \Omega_t \mathbf{C}\, dw = -\int_{\phi} \Omega \mathbf{C}(-\lambda) \phi_t\, ds + \int_{G_2} \Omega \mathbf{C}_t\, dw.$$

We now add the left sides and right sides of equations (15), (16), (17), (18). Equations (13) and (12) tell us that the left side of the sum is zero. On the right side of the sum the two integrals over G_1 add up to zero because $\mathbf{E}$ and $\mathbf{H}$ are required to satisfy Maxwell's equations in G_1, and for the same reason the two integrals over G_2 add to zero. If we now combine the remaining integrals over ϕ using the symbol [ ] to denote the jump discontinuity of the enclosed quantity over $\phi = 0$, we obtain

$$(19) \qquad \int_{\phi} \Omega(\text{grad}\, \phi \times [\mathbf{H}] - \phi_t[\mathbf{C}])\lambda\, ds = 0.$$

Similarly we obtain from equation (12b) that

$$(20) \qquad \int_{\phi} \Omega(\text{grad}\, \phi \times [\mathbf{E}] + \frac{1}{c} \phi_t[\mu \mathbf{H}])\lambda\, ds = 0.$$

Since these integrals must be zero for any choice of Ω (which vanishes on the boundary and outside of G) the other factor in each integrand must be zero at each point of ϕ. We obtain therefore

$$(21) \qquad \text{grad}\, \phi \times [\mathbf{H}] - \frac{1}{c} \phi_t[\epsilon \mathbf{E} + \mathbf{F}] = 0,$$

$$(22) \qquad \text{grad}\, \phi \times [E] + \frac{1}{c} \phi_t[\mu \mathbf{H}] = 0,$$

which must hold on the discontinuity surface $\phi = 0$.[11]

[We shall now use conditions (21) and (22) to dispose of a point left un-

---

[11]These same discontinuity conditions are arrived at by a different method in R. K. Luneberg, *Mathematical Theory of Optics*, Brown University, 1944, p. 21. They yield as special cases the usual discontinuity conditions on $\mathbf{E}$ and $\mathbf{H}$ which hold on the boundary $\psi(x, y, z) = 0$ separating media of different electromagnetic properties. Also, since $\phi \equiv t = 0$ is a discontinuity surface, (21) yields $\mathbf{E}(0+) = -\mathbf{g}\, f(0+)$. But for the pulse solution $f(0+) = 1$. Hence $\mathbf{E}_0\, (0+) = -\mathbf{g}/\epsilon$. From equation (22) we obtain $\mathbf{H}(0+) = 0$.

settled in section 2 (cf. footnote 2). As pointed out there, we know from Maxwell's equations that

$$\text{(a)} \qquad \frac{\partial}{\partial t}(\operatorname{div} \epsilon\mathbf{E} + \operatorname{div} \mathbf{F}) = 0.$$

The hypersurface $\phi \equiv t = 0$ is a discontinuity surface for **E**, **H**, and **F**. For this ϕ, grad $\phi = 0$ and $\phi_t \neq 0$. Hence equation (21) tells us that $[\epsilon\mathbf{E} + \mathbf{F}] = 0$ at $t = 0$. Since **E** and **F** are 0 for $t < 0$ we know that $\epsilon\mathbf{E} + \mathbf{F} = 0$ for $t = 0+$. Since ϵ, **E**, and **F** are sectionally continuous, $\epsilon\mathbf{E} = -\mathbf{F}$ for some t-interval about $t = 0+$ and in this t-interval div $\epsilon\mathbf{E} = -$div **F**, that is, no additional $f(x, y, z)$ is needed to integrate (a) with respect to t.

We now prove that div $\epsilon\mathbf{E} = -$div **F** for all t. Let t_1 be a value of t at which **E** or **F** is discontinuous and suppose div $\epsilon\mathbf{E} = -$div **F** for $t < t_1$. For $t > t_1$, the relation div $\epsilon\mathbf{E} = -$ div **F** need not hold for a jump in the value of either $\epsilon\mathbf{E}$ or **F** may introduce an $f(x, y, z)$. However, taking the divergence of (21) and using vector identities gives

$$\text{(b)} \qquad -\operatorname{grad} \phi \cdot \operatorname{curl} [\mathbf{H}] - \frac{\phi_t}{c} \operatorname{div} [\epsilon\mathbf{E} + \mathbf{F}] - [\epsilon\mathbf{E} + \mathbf{F}] \cdot \operatorname{grad} \frac{\phi_t}{c} = 0.$$

Since Maxwell's equation (1a) holds on either side of $t = t_1$ we may write

$$\text{(c)} \qquad \operatorname{curl} [\mathbf{H}] - \frac{1}{c} [\epsilon\mathbf{E}_t + \mathbf{F}_t] = 0.$$

We substitute for curl [**H**] in (b) and obtain

$$\text{(d)} \qquad -\operatorname{grad} \phi \cdot \frac{1}{c} [\epsilon\mathbf{E}_t + \mathbf{F}_t] - \frac{\phi_t}{c} \operatorname{div} [\epsilon\mathbf{E} + \mathbf{F}] - [\epsilon\mathbf{E} + \mathbf{F}] \cdot \operatorname{grad} \frac{\phi_t}{c} = 0.$$

or

$$\text{(e)} \qquad -\frac{\partial}{\partial t}\left(\operatorname{grad} \phi \cdot \frac{1}{c} [\epsilon\mathbf{E} + \mathbf{F}]\right) - \frac{\phi_t}{c} \operatorname{div} [\epsilon\mathbf{E} + \mathbf{F}] = 0$$

If we dot (21) by grad ϕ we see that the first term in (e) is zero. Now if $\phi = 0$ is a hypersurface for which $\phi_t \neq 0$ at $t = t_1$, then [div $\epsilon\mathbf{E}$ + div **F**] $= 0$ on $\phi = 0$ which means again, as in the case $\phi \equiv t = 0$, that no $f(x, y, z)$ can arise by integrating (a) with respect to t.]

Returning to the discontinuity conditions (21) and (22), we see that if ϵ, μ and **F** are continuous at a point (x, y, z, t) on $\phi = 0$ and if $\phi = 0$ can be represented in the form $\psi - ct = 0$ then equations (21) and (22) specialize to

$$\text{(23)} \qquad \operatorname{grad} \psi \times [\mathbf{H}] + \epsilon[\mathbf{E}] = 0,$$

$$\text{(24)} \qquad \operatorname{grad} \psi \times [\mathbf{E}] - \mu[\mathbf{H}] = 0$$

on the surface $\psi - ct = 0$.

Equations (23) and (24) are six homogeneous equations in the six quantities, namely, the components of [E] and [H]. Since discontinuities are assumed to exist on $\psi - ct = 0$, [E] and [H] are not both zero. Hence the determinant of the coefficients must vanish. This condition is both necessary and sufficient. However, it is also possible to show[12] that a necessary and sufficient condition for non-zero solutions of the six equations is

$$\psi_x^2 + \psi_y^2 + \psi_z^2 = \epsilon\mu. \tag{25}$$

This equation is the eiconal equation of geometrical optics and it means that $\psi - ct = 0$ is a family of wave-fronts. Thus at a point in x, y, z, t space where ϵ, μ, and **F** are continuous, but **E** and **H** discontinuous, a discontinuity surface must be a wave front.

We shall now derive conditions which the discontinuities of $\mathbf{E}_t$ and $\mathbf{H}_t$ must satisfy on a discontinuity surface $\phi = 0$. Since the Ω which was used to derive equations (12) is arbitrary (except for the condition on the boundary and exterior of G) we may replace it by Ω_t which must also vanish on the boundary and exterior of G. We may therefore state instead of (12a)

$$\int_G (\text{grad } \Omega_t \times \mathbf{H} - \Omega_{tt}\mathbf{C})\, dw = 0. \tag{26}$$

We shall now use the fact of equation (13), namely, that the integral over G is equal to the integral over G_1 plus the integral over G_2. We consider the integral over G_1 and apply integration by parts with respect to t, obtaining

$$\begin{aligned} \int_{G_1} (\text{grad } \Omega_t \times \mathbf{H} - \Omega_{tt}\mathbf{C})\, dw &= \int_{\Gamma_1+\phi} (\text{grad } \Omega \times \mathbf{H} - \Omega_t\mathbf{C})\, dx\, dy\, dz \\ &\quad - \int_{G_1} (\text{grad } \Omega \times \mathbf{H}_t - \Omega_t\mathbf{C}_t)\, dw. \end{aligned} \tag{27}$$

Because Ω and Ω_t are zero on $\Gamma = \Gamma_1 + \Gamma_2$, the "surface" integral over Γ_1 vanishes. However, the integral over the portion of $\phi = 0$ contained in G need not vanish.

We now consider the term

$$\int_\phi \text{grad } \Omega \times \mathbf{H}\, dx\, dy\, dz. \tag{28}$$

In Ω and **H** we must understand that the variable t is replaced by $\psi(x, y, z)/c$ which is the form of $\phi = 0$ when solved for t. However, the gradient operator

[12]Replace [H] in (23) by its value from (24). If **E** and **H** are continuous on $\psi - ct = 0$ then the condition that ψ satisfy the eiconal equation need not hold. In this case the condition would follow from the next set of discontinuity conditions, namely, equations (39a) and (39b), if we use the fact that the right sides are zero for continuous **E** and **H**.

applied to Ω applies only to the x, y, z of $\Omega(x, y, z, t)$. Let us denote by $\partial\Omega/\partial y$ the derivative of $\Omega(x, y, z, \psi(x, y, z)/c)$, whereas Ω_y denotes the partial derivative of $\Omega(x, y, z, t)$. Then

$$\frac{\partial\Omega}{\partial y} = \Omega_y + \frac{1}{c}\,\Omega_t\psi_y\,. \tag{29}$$

For a typical term of (28), such as $\int_\phi \Omega_y H_3\,dx\,dy\,dz$, we may say

$$\int_\phi \Omega_y H_3\,dx\,dy\,dz = \int_\phi \frac{\partial\Omega}{\partial y} H_3\,dx\,dy\,dz - \frac{1}{c}\int_\phi \Omega_t\psi_y H_3\,dx\,dy\,dz. \tag{30}$$

If we integrate by parts the first term on the right with respect to y we obtain

$$\int_\phi \frac{\partial\Omega}{\partial y} H_3\,dx\,dy\,dz = \int_{\text{Boundary of }\phi} \Omega H_3\,dx\,dz - \int_\phi \Omega\frac{\partial H_3}{\partial y}\,dx\,dy\,dz. \tag{31}$$

In the first integral on the right side we must substitute in Ω and H_3 the value of y on the boundary of the portion of ϕ which lies in G. Since Ω is 0 on this boundary this integral vanishes.

As to the second term on the right side of (30), a theorem on differentiation of an implicit function, applied to $\phi(x, y, z, t) = 0$, enables us to say that

$$\frac{1}{c}\int_\phi \Omega_t\psi_y H_3\,dx\,dy\,dz = \int_\phi \Omega_t H_3\left(\frac{-\phi_y}{\phi_t}\right)dx\,dy\,dz. \tag{32}$$

Hence using the facts of (31) and (32) in (30) and remembering that (30) is merely a typical term of (28) we may say that

$$\begin{aligned}\int_\phi \operatorname{grad}\Omega\times\mathbf{H}\,dx\,dy\,dz &= -\int_\phi \Omega\operatorname{curl}\mathbf{H}\left(x, y, z, \frac{1}{c}\,\psi(x, y, z)\right)dx\,dy\,dz\\ &\quad+\int_\phi \Omega_t\,\frac{\operatorname{grad}\phi\times\mathbf{H}}{\phi_t}\,dx\,dy\,dz.\end{aligned} \tag{33}$$

In (33) the gradient operator applies only to the x, y, z which appear before t is replaced by ψ/c. However, the curl operator applies to all the x, y, z.

We may replace the volume element $dx\,dy\,dz$ by $\lambda\phi_t\,ds$ and (33) becomes

$$\begin{aligned}&\int_\phi \operatorname{grad}\Omega\times\mathbf{H}\lambda\phi_t\,ds\\ &\quad= -\int_\phi \Omega\operatorname{curl}\mathbf{H}\left(x, y, z, \frac{\psi}{c}\right)\lambda\phi_t\,ds + \int_\phi \Omega_t\operatorname{grad}\phi\times\mathbf{H}\lambda\,ds.\end{aligned} \tag{34}$$

We now return to (27) and replace $dx\,dy\,dz$ there by $\lambda\phi_t\,ds$. We therefore have from (34) and (27)

$$\int_{G_1} (\text{grad } \Omega_t \times \mathbf{H} - \Omega_{tt}\mathbf{C})\, dw$$

$$(35) \qquad = \int_\phi \{-\Omega \text{ curl } \mathbf{H}\phi_t + \Omega_t \text{ grad } \phi \times \mathbf{H} - \Omega_t\phi_t\mathbf{C}\}\lambda\, ds - \int_{G_1} (\text{grad } \Omega \times \mathbf{H}_t - \Omega_t\mathbf{C}_t)\, dw.$$

We may now apply steps (15) and (17) to the volume integral on the right side of (35) except that $\mathbf{H}_t$ and $\mathbf{C}_t$ here replace $\mathbf{H}$ and $\mathbf{C}$ there. We obtain that

$$\int_{G_1} (\text{grad } \Omega_t \times \mathbf{H} - \Omega_{tt}\mathbf{C})\, dw$$

$$(36) \qquad = \int_\phi \{-\Omega \text{ curl } \mathbf{H}\phi_t + \Omega_t \text{ grad } \phi \times \mathbf{H} - \Omega_t\phi_t\mathbf{C}\}\lambda\, ds - \int_\phi \{\Omega \text{ grad } \phi \times \mathbf{H}_t - \Omega\phi_t\mathbf{C}_t\}\lambda\, ds + \int_{G_1} (\Omega \text{ curl } \mathbf{H}_t - \Omega\mathbf{C}_{tt})\, dw.$$

Of course the same statement can be made for G_2 the only change being that the normal has the opposite direction over ϕ. Also if we take the time derivative of Maxwell's equation (1a), multiply by Ω, and integrate over G_1, we see that the volume integral on the right side of (36) is zero because one factor of the integrand is zero in G_1. Likewise the same integral over G_2 is zero. We now add (36) to the same equation applied to G_2. We use the remarks just made and (26). Then

$$(37) \qquad \int_\phi \left(-\Omega \text{ curl}\left[\mathbf{H}\left(x, y, z, \frac{\psi}{c}\right)\right]\phi_t + \Omega_t \text{ grad } \phi \times [\mathbf{H}] - \Omega_t[\mathbf{C}]\phi_t\right)\lambda\, ds = +\int_\phi (\Omega \text{ grad } \phi \times [\mathbf{H}_t] - \Omega[\mathbf{C}_t]\phi_t)\lambda\, ds.$$

If we multiply (21) by Ω_t we see that the second and third terms on the left side of (37) vanish.

Again we have an integral over ϕ with Ω arbitrary. Hence the other factor of the integrand must vanish and we obtain

$$(38a) \qquad \text{grad } \phi \times [\mathbf{H}_t] - \frac{\phi_t}{c}[\epsilon\mathbf{E}_t + \mathbf{F}_t] = -\phi_t \text{ curl}\left[\mathbf{H}\left(x, y, z, \frac{\psi}{c}\right)\right].$$ [13]

[13] We note that at the discontinuity surface $\phi \equiv t = 0$, (38a) reduces to $-(\epsilon/c)E_t(0+) - (1/c)F_t(0+) = -\text{curl } H(x, y, z, 0+)$. We showed in footnote 11 that $H(0+) = 0$. Using the continuity of the derivatives of H in and on the boundary (from one side of the discontinuity surface) of G_1, or G_2, we conclude that curl $H(0+) = 0$. For the pulse solution $F_t(0+) = 0$ and hence $E_{0t}(0+) = 0$. We may argue similarly from (38b) to conclude that $H_{0t}(0+) = 0$ except at points $(x, y, z, 0)$ where charges exist.

Likewise we obtain from equation (12b), by steps analogous to steps (26) to (38a),

$$\text{(38b)} \qquad \operatorname{grad} \phi \times [\mathbf{E}_t] + \frac{1}{c}\,\phi_t[\mu \mathbf{H}_t] = -\phi_t \operatorname{curl}\left[\mathbf{E}\left(x, y, z, \frac{\psi}{c}\right)\right].$$

If ϵ, μ, and $\mathbf{F}$ are continuous at x, y, z, t on $\phi = 0$ and if ϕ may be written in the form $\psi - ct$, then

$$\text{(39a)} \qquad \operatorname{grad} \psi \times [\mathbf{H}_t] + \epsilon[\mathbf{E}_t] = c \operatorname{curl} [\mathbf{H}(x, y, z, \psi/c)],$$

$$\text{(39b)} \qquad \operatorname{grad} \psi \times [\mathbf{E}_t] - \mu[\mathbf{H}_t] = c \operatorname{curl} [\mathbf{E}(x, y, z, \psi/c)].$$

Relations (38) and (39) hold on the discontinuity surface $\phi = 0$ and $\psi = ct$, respectively.[14]

The process used to derive equations (38), may be applied once more to obtain conditions on $[\mathbf{E}_{tt}]$ and $[\mathbf{H}_{tt}]$. One starts with the equations (12) wherein since Ω is arbitrary we may replace it by Ω_{tt}. We use the fact that the integral over G equals the sum of the integral over G_1 and the integral over G_2. Consider first the integral over G_1, namely,

$$\int G_1(\operatorname{grad} \Omega_{tt} \times \mathbf{H} - \Omega_{ttt}\mathbf{C})\, dw.$$

We integrate with respect to t and obtain (as in step (27))

$$\int_\phi (\operatorname{grad} \Omega_t \times \mathbf{H} - \Omega_{tt}\mathbf{C})\, dx\, dy\, dz - \int_{G_1} (\operatorname{grad} \Omega_t \times \mathbf{H}_t - \Omega_{tt}\mathbf{C}_t)\, dw.$$

We may now operate on the first integral as in the steps from (27) to (35). In fact the only change is that Ω_t replaces Ω and Ω_{tt} replaces Ω_t. This gives, in view of (35),

$$\int_\phi (-\Omega_t \operatorname{curl} \mathbf{H}\phi_t + \Omega_{tt} \operatorname{grad} \phi \times \mathbf{H} - \Omega_{tt}\mathbf{C}\phi_t)\lambda\, ds$$

$$- \int_{G_1} (\operatorname{grad} \Omega_t \times \mathbf{H}_t - \Omega_{tt}\mathbf{C}_t)\, dw.$$

---

[14] Conditions (38) and (39) are non-homogeneous while conditions (21) to (24) are homogeneous. This result seems surprising since $\mathbf{E}_t$ and $\mathbf{H}_t$ are solutions of Maxwell's equations just as $\mathbf{E}$ and $\mathbf{H}$ are; moreover, $\mathbf{E}_t$ and $\mathbf{H}_t$ satisfy definite initial conditions as do $\mathbf{E}$ and $\mathbf{H}$. However, we have supposed that $\mathbf{E}$ and $\mathbf{H}$, if discontinuous, are limits of continuous solutions and this condition need not be satisfied by $\mathbf{E}_t$ and $\mathbf{H}_t$. In special cases it could however be that $\mathbf{E}_t$ and $\mathbf{H}_t$ are limits of continuous solutions, in which cases conditions (38) and (39) would be homogeneous. But the discontinuity conditions on higher time derivatives of $\mathbf{E}$ and $\mathbf{H}$ may then still be non-homogeneous.

In view of (21) we shall be able to throw out the last two terms of the first integral after we have combined results for G_1 and G_2. We therefore neglect them now. We next treat the volume integral. Again apply the steps analogous to those from (27) to (35) except that $\mathbf{H}_t$ replaces $\mathbf{H}$ and $\mathbf{C}_t$ replaces $\mathbf{C}$. This operation gives

$$\int_\phi - \Omega_t \operatorname{curl} \mathbf{H}\lambda\phi_t \, ds - \int_\phi (-\Omega \operatorname{curl} \mathbf{H}_t\phi_t + \Omega_t \operatorname{grad} \phi \times \mathbf{H}_t - \Omega_t\mathbf{C}_t\phi_t)\lambda \, ds$$

$$+ \int_{G_1} (\operatorname{grad} \Omega \times \mathbf{H}_{tt} - \Omega_t\mathbf{C}_{tt}) \, dw.$$

In view of (38a) we see that we shall be able to throw out those terms containing Ω_t after we have combined results for G_1 and G_2. We therefore do so now and are left with

$$\int_\phi \Omega \operatorname{curl} \mathbf{H}_t\phi_t\lambda \, ds + \int_{G_1} (\operatorname{grad} \Omega \times \mathbf{H}_{tt} - \Omega_t\mathbf{C}_{tt}) \, dw.$$

We now operate on the volume integral as in the steps (15) and (17) and obtain

$$\int_\phi \Omega \operatorname{curl} \mathbf{H}_t\phi_t\lambda \, ds + \int_\phi (\Omega \operatorname{grad} \phi \times \mathbf{H}_{tt} - \Omega\mathbf{C}_{tt}\phi_t)\lambda \, ds$$

$$- \int_{G_1} (\Omega \operatorname{curl} \mathbf{H}_{tt} - \Omega\mathbf{C}_{ttt}) \, dw.$$

Again we may throw out the volume integral since it follows from Maxwell's equation (1a) that in a domain G_1, in which $\mathbf{H}_{tt}$ and $\mathbf{C}_{ttt}$ exist, the volume integral is zero. The same result applies to G_2 and adding as in step (19) or (37) we get the next discontinuity condition

$$\operatorname{grad} \phi \times [\mathbf{H}_{tt}] - \frac{1}{c}\phi_t[\epsilon\mathbf{E}_{tt} + \mathbf{F}_{tt}] = -\phi_t \operatorname{curl}\left[\mathbf{H}_t\left(x, y, z, \frac{\psi}{c}\right)\right].$$

Similarly, we obtain

$$\operatorname{grad} \phi \times [\mathbf{E}_{tt}] + \frac{1}{c}\phi_t[\mu\mathbf{H}_{tt}] = -\phi_t \operatorname{curl}\left[\mathbf{E}_t\left(x, y, z, \frac{\psi}{c}\right)\right].$$

Again if $\phi = \psi - ct$ and if ϵ, μ, and $\mathbf{F}$ are continuous on $\psi - ct$ then

(40a) $$\operatorname{grad} \psi \times [\mathbf{H}_{tt}] + \epsilon[\mathbf{E}_{tt}] = c \operatorname{curl} [\mathbf{H}_t];$$

(40b) $$\operatorname{grad} \psi \times [\mathbf{E}_{tt}] - \mu[\mathbf{H}_{tt}] = c \operatorname{curl} [\mathbf{E}_t].$$

In these last two sets of formulas the curl operation acts on all x, y, z in

$\mathbf{H}_t(x, y, z, \psi/c)$ and $\mathbf{E}_t(x, y, z, \psi/c)$ but the gradient acts only on the x, y, z in ϕ which are present before substituting ψ/c for t.

We now change notation so that

$$\mathbf{A}_\nu = \left[\frac{\partial^\nu \mathbf{E}}{\partial t^\nu}\right] \quad \text{and} \quad \mathbf{B}_\nu = \left[\frac{\partial^\nu \mathbf{H}}{\partial t^\nu}\right] \quad \text{for} \quad \nu \geq 0$$

and $\mathbf{A}_{-1} = \mathbf{B}_{-1} = 0$. The process we used to obtain equations (23) and (24), (39), and (40) obviously generalizes and we may write for $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ and $\nu \geq 0$

$$\text{(41a)} \qquad \operatorname{grad} \psi \times \mathbf{B}_\nu + \epsilon \mathbf{A}_\nu = c \operatorname{curl} \mathbf{B}_{\nu-1}(x, y, z, \psi/c),$$

$$\text{(41b)} \qquad \operatorname{grad} \psi \times \mathbf{A}_\nu - \mu \mathbf{B}_\nu = c \operatorname{curl} \mathbf{A}_{\nu-1}(x, y, z, \psi/c).$$

Relations (41) hold on the discontinuity surface $\psi = ct$.[15]

We have six linear equations for the components of $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$, the very quantities needed for the coefficients of the asymptotic expansion of the steady state fields. Unfortunately the determinant of the coefficients of these linear equations is zero, a fact noted in connection with equations (23) and (24). Since non-zero solutions of equations (41) are known to exist in the case of a pulse source say, it follows that the right sides of (41) must satisfy some additional conditions.[16] We proceed to derive these conditions.

---

[15]By starting with Maxwell's equation and by assuming that the time behavior of the solutions is $\exp\{-i\omega t\}$ so that $\mathbf{E} = \mathbf{u} \exp\{-i\omega t\}$, $\mathbf{H} = \mathbf{v} \exp\{-i\omega t\}$, one obtains the time-free form of Maxwell's equations involving the functions $\mathbf{u}$ and $\mathbf{v}$. If we now assume that

$$\mathbf{u} = \sum_0^\infty \mathbf{U}_\nu \left(\frac{1}{ik}\right)^\nu e^{ik\psi} \quad \text{and} \quad \mathbf{v} = \sum_0^\infty \mathbf{V}_\nu \left(\frac{1}{ik}\right)^\nu e^{ik\psi},$$

substitutes in Maxwell's equations, and equates to zero the coefficients of like powers of $1/ik$, we obtain conditions (41). (See Luneberg, *Mathematical Theory of Optics*, Brown University Notes, 1944, Volume I, p. 81.) This process assumes the existence and differentiability of the asymptotic series for $\mathbf{u}$ and $\mathbf{v}$. (Existence is established in this paper at least for a class of solutions of Maxwell's equations.) Further, the process just indicated gives no inkling of the relationship of the $\mathbf{U}_\nu$ and $\mathbf{V}_\nu$ to the discontinuities of the pulse solution $\mathbf{E}_0$, $\mathbf{H}_0$. One could by an indirect argument which relies upon the uniqueness of asymptotic expansions subsequently identify the $\mathbf{U}_\nu$ and $\mathbf{V}_\nu$ with these discontinuities. However, were the substitute procedure presented carefully it might at best be slightly briefer at the expense of directness and insight. Incidentally, conditions (41) as given by Luneberg have a minus sign on the right side but this difference is due to the difference in signs between the asymptotic expansion (6) of this paper and the expansion assumed by Luneberg in the Brown notes.

[16]At first sight one might expect to solve equations (41) by a purely algebraic procedure, since these are just six linear equations in the six components of $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$. However, it can be shown that the rank of the matrix of the coefficients is four. Hence even if one proceeded to solve, relying upon the fact that non-zero solutions are known to exist, the solution can at best express four of the components in terms of the other two. The solutions obtained for any one value of ν must be chosen so that, when used on the right side of equations (41), it would be possible to solve for the $\mathbf{A}_{\nu+1}$ and $\mathbf{B}_{\nu+1}$; in other words, the $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ must be chosen so that

From equation (41a)

$$\mathbf{A}_\nu = -\frac{1}{\epsilon}\operatorname{grad}\psi \times \mathbf{B}_\nu + \frac{c}{\epsilon}\operatorname{curl}\mathbf{B}_{\nu-1}\ .$$

We substitute this quantity in (41b) and obtain

$$-\operatorname{grad}\psi \times \left(\frac{1}{\epsilon}\operatorname{grad}\psi \times \mathbf{B}_\nu\right) + \operatorname{grad}\psi \times \frac{c}{\epsilon}\operatorname{curl}\mathbf{B}_{\nu-1} - \mu\mathbf{B}_\nu = c\operatorname{curl}\mathbf{A}_{\nu-1}\ .$$

Applying the vector identity for the vector triple product gives

$$\begin{aligned} &-\frac{1}{\epsilon}(\operatorname{grad}\psi\cdot\mathbf{B}_\nu)\operatorname{grad}\psi + \frac{1}{\epsilon}(\operatorname{grad}\psi)^2\mathbf{B}_\nu - \mu\mathbf{B}_\nu \\ &= -\frac{c}{\epsilon}\operatorname{grad}\psi \times \operatorname{curl}\mathbf{B}_{\nu-1} + c\operatorname{curl}\mathbf{A}_{\nu-1}\ . \end{aligned} \tag{43}$$

We form the scalar product of both sides of this equation with $\mathbf{A}_0$ and use the fact [17] that $\mathbf{A}_0\cdot\operatorname{grad}\psi = 0$. We obtain

$$\frac{1}{\epsilon}(\operatorname{grad}\psi)\mathbf{A}_0\cdot\mathbf{B}_\nu - \mathbf{A}_0\cdot\mathbf{B}_\nu = -\frac{c}{\epsilon}\mathbf{A}_0\cdot\operatorname{grad}\psi \times \operatorname{curl}\mathbf{B}_{\nu-1} + c\mathbf{A}_0\cdot\operatorname{curl}\mathbf{A}_{\nu-1}\ .$$

By equation (24) $\mathbf{A}_0 \times \operatorname{grad}\psi = -\mu\mathbf{B}_0$. Hence, first interchanging dot and cross on the right side, we obtain

$$\frac{1}{\epsilon}(\operatorname{grad}\psi)^2\mathbf{A}_0\cdot\mathbf{B}_\nu - \mu\mathbf{A}_0\cdot\mathbf{B}_\nu = +\frac{c}{\epsilon}\mu\mathbf{B}_0\cdot\operatorname{curl}\mathbf{B}_{\nu-1} + c\mathbf{A}_0\cdot\operatorname{curl}\mathbf{A}_{\nu-1}\ .$$

In view of equation (25) we obtain

$$\mu\mathbf{B}_0\cdot\operatorname{curl}\mathbf{B}_{\nu-1} + \epsilon\mathbf{A}_0\cdot\operatorname{curl}\mathbf{A}_{\nu-1} = 0. \tag{44a}$$

If we form the dot product of both sides of (43) with $\mathbf{B}_0$ and use the fact[17] that $\mathbf{B}_0\cdot\operatorname{grad}\psi = 0$ we get

$$\left(\frac{1}{\epsilon}(\operatorname{grad}\psi)^2\mathbf{B}_\nu - \mu\mathbf{B}_\nu\right)\cdot\mathbf{B}_0 = -\frac{c}{\epsilon}\mathbf{B}_0\cdot\operatorname{grad}\psi \times \operatorname{curl}\mathbf{B}_{\nu-1} + c\mathbf{B}_0\cdot\operatorname{curl}\mathbf{A}_{\nu-1}\ .$$

---

the next set of equations in the recursive system (41) is consistent. But these consistency conditions are precisely the equations (45) which we derive here and unfortunately these conditions are partial differential equations. Hence we are led, as in the procedure of the paper, to solve a mixed system of algebraic and partial differential equations.

[17] This fact follows at once from equation (23) if we form the scalar product of grad ψ with the left side. Likewise we get from (24) that $B_0\cdot\operatorname{grad}\psi = 0$.

If we interchange dot and cross and use (23) and (25) we get

$$\text{(44b)} \qquad \mathbf{A}_0 \cdot \operatorname{curl} \mathbf{B}_{\nu-1} - \mathbf{B}_0 \cdot \operatorname{curl} \mathbf{A}_{\nu-1} = 0.$$

Equations (44) are the conditions which hold for the right sides of equations (41) in view of the fact that non-zero solutions must exist for $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$. Since equations (41) hold for each value of $\nu \geq 0$, we can state equations (44) for ν instead of $\nu - 1$, namely,

$$\text{(45)} \qquad \begin{aligned} \mu \mathbf{B}_0 \cdot \operatorname{curl} \mathbf{B}_\nu + \epsilon \mathbf{A}_0 \cdot \operatorname{curl} \mathbf{A}_\nu &= 0, \\ \mathbf{A}_0 \cdot \operatorname{curl} \mathbf{B}_\nu - \mathbf{B}_0 \cdot \operatorname{curl} \mathbf{A}_\nu &= 0; \end{aligned} \qquad \nu \geq 0.$$

Equations (41) supplemented by equations (45) are the discontinuity conditions which relate solutions of Maxwell's differential equations on the two sides of a discontinuity surface $\psi - ct = 0$. It is evident that these conditions are consequences of the integral formulation of Maxwell's equations.

5. *The Ordinary Differential Equations for the Discontinuities* $\mathbf{A}_\nu$ *and* $\mathbf{B}_\nu$

We shall use equations (41) and (45) to determine $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$. While equations (41) are just ordinary algebraic equations, equations (45) are first order partial differential equations. Instead of attempting to solve these partial differential equations directly we shall convert them, with the aid of equations (41), to a system of recursive ordinary differential equations. The process which will be used generalizes the one employed[18] to obtain the ordinary differential equations for $\mathbf{A}_0$ and $\mathbf{B}_0$ which hold along the rays, that is, the curves which are orthogonal to the wave fronts $\psi - ct = 0$. The latter differential equations are known as the transport equations of geometrical optics. The additional ordinary differential equations which we shall obtain and which will give the variation of $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ along the rays may properly be called the higher transport equations.

Because the derivation of the ordinary differential equations is lengthy it has been treated separately in the appendix. The results[19] are:

---

[18] See Luneberg, Brown Notes, Article 11, pp. 41–46. See in particular equations (11.38) of this reference.

[19] Equations (46a) and (46b) here are equations (48) and (49) of the Appendix. It was remarked in footnote 15 that the discontinuity conditions (41) can be derived by assuming the existence and differentiability of an asymptotic solution to Maxwell's equations and by substituting in the equations. It is worthy of note that equations (46) can be obtained formally by substitution of the asymptotic solution in the second order equations for **E** and **H** derived from Maxwell's equations. This fact has been noted by H. J. Riblet in an unpublished paper and by F. G. Friedlander in "Geometrical Optics and Maxwell's Equations," *Proc. Cambridge Phil. Soc.*, Vol. 43, Part 2, April, 1947, pp. 284–286. The comments made in footnote 15 a propos of the procedure discussed there apply to the procedure mentioned here.

$$(46a)\qquad 2\frac{d\mathbf{A}_\nu}{d\tau} + \mathbf{A}_\nu \Delta_\mu \psi + \frac{2}{n}(\text{grad}\, n \cdot \mathbf{A}_\nu)\, \text{grad}\, \psi = -\mathbf{C}_\nu$$

$$(46b)\qquad 2\frac{d\mathbf{B}_\nu}{d\tau} + \mathbf{B}_\nu \Delta_\epsilon \psi + \frac{2}{n}(\text{grad}\, n \cdot \mathbf{B}_\nu)\, \text{grad}\, \psi = -\mathbf{D}_\nu\,,$$

wherein

$$\mathbf{C}_\nu = \mu\, \text{curl}\left(\frac{c}{\mu}\, \text{curl}\, \mathbf{A}_{\nu-1}\right) - \text{grad}\left(\frac{c}{\epsilon}\, \text{div}\, \epsilon \mathbf{A}_{\nu-1}\right),$$

$$\mathbf{D}_\nu = \epsilon\, \text{curl}\left(\frac{c}{\epsilon}\, \text{curl}\, \mathbf{B}_{\nu-1}\right) - \text{grad}\left(\frac{c}{\mu}\, \text{div}\, \mu \mathbf{B}_{\nu-1}\right),$$

$$\mathbf{C}_0 = \mathbf{D}_0 = 0,$$

$$\Delta_\mu \psi = \mu\left(\left(\frac{\psi_x}{\mu}\right)_x + \left(\frac{\psi_y}{\mu}\right)_y + \left(\frac{\psi_z}{\mu}\right)_z\right),$$

and

$$\Delta_\epsilon \psi = \epsilon\left(\left(\frac{\psi_x}{\epsilon}\right)_x + \left(\frac{\psi_y}{\epsilon}\right)_y + \left(\frac{\psi_z}{\epsilon}\right)_z\right).$$

It is further understood that equations (46) represent differentiation along the rays orthogonal to the wave fronts $\psi - ct = 0$, τ being any convenient parameter in terms of which the equations of the rays can be expressed.

To use equations (46) we note first of all that they are a series of recursive vector differential equations for the six quantities, the components of $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$. Thus $\mathbf{A}_0$ and $\mathbf{B}_0$ must be determined first and then used on the right side to solve the next case for $\nu = 1$, etc.

We note next that one needs the quantity $\psi(x, y, z)$ to fix the coefficients in (46). Now ψ satisfies the eiconal equation $\psi_x^2 + \psi_y^2 + \psi_z^2 = \epsilon\mu$ and hence one must solve this partial differential equation. In the simple case of a dipole in a non-homogeneous medium there will be a single set of expanding wave fronts. If, however, there are obstacles or other discontinuities in the medium there will be wave-fronts reflected from the discontinuities. Not only must these wave fronts be determined but the differential equations (46) must be solved along the rays to these wave fronts also. The existence of more than one set of wave fronts means that at any point x, y, z of space more than one discontinuity surface $\psi(x, y, z) = ct$ will pass and the several values of $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ at this point will be required to obtain the asymptotic expansion (6). That is, a summation over α will have to be applied, the various $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ at one point being actually then the several values which must be summed.

A third point to note in regard to the use of equations (46) is that we want the known coefficients to be expressed as functions of τ, the parameter along the rays. This means that the quantities $\Delta_\mu\psi$, grad n, n, and grad ψ

must be expressed as functions of τ. Theoretically this can be done for, from a knowledge of the wave fronts one can obtain the differential equations for the rays and then solve these equations for the rays.[20] Once we obtain the equations of the wave fronts and the rays we can at least fix the coefficients on the left side of (46). We must still determine the quantities on the right side before solving for $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$. Now the quantities $\mathbf{C}_\nu$ and $\mathbf{D}_\nu$ will be known to us as functions of τ from the solution of the preceding differential equation in the recursive set. In order to determine such quantities as curl $\mathbf{A}_{\nu-1}$ and div $\mathbf{A}_{\nu-1}$ we need $\mathbf{A}_{\nu-1}$ as a function of x, y, z. For this purpose we may invert the equations of the rays themselves. To each family of wavefronts there corresponds a two parameter family of rays. These two parameters will be involved in our functions $\mathbf{A}_{\nu-1}$ and $\mathbf{B}_{\nu-1}$ because we substituted the equations of the rays into ψ, n, and other quantities in order to determine $\mathbf{A}_{\nu-1}$ and $\mathbf{B}_{\nu-1}$. We may now take the equations of the rays

$$x = x(\tau, \alpha, \beta), \qquad y = y(\tau, \alpha, \beta), \qquad s = s(\tau, \alpha, \beta) \tag{47}$$

and solve these for α, β, τ in terms of x, y, and z. Substitution of these values of α, β, and τ in $\mathbf{A}_{\nu-1}(\tau, \alpha, \beta)$ and $\mathbf{B}_{\nu-1}(\tau, \alpha, \beta)$ converts them into functions of x, y, and z.

The solutions of equations (46) for the case $\nu = 0$ (in which case $\mathbf{C}_\nu = \mathbf{D}_\nu = 0$) give the geometrical optics behavior of the field generated by a source along any particular ray emanating from that source, for expansions (6) show that, except for a phase factor, $\mathbf{A}_0$ and $\mathbf{B}_0$ give the field generated by a monochromatic source as the wave length approaches zero.[21] While the equations for $\mathbf{A}_0$ and $\mathbf{B}_0$ have been derived by other methods it is the asymptotic expansions (6) which show clearly what they furnish.

6. *Introduction of Initial or Boundary Conditions for* $\mathbf{A}_\nu$ *and* $\mathbf{B}_\nu$

The discussion of equations (46) in the preceding section ignored the fact that solution of these equations introduces arbitrary constants. Hence thus far $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ are not uniquely determined. In fact the differential equations (45) and linear equations (41), as well as the resulting ordinary differential equations (46), hold for all solutions $\mathbf{E}$ and $\mathbf{H}$ of the integral form (12) of Maxwell's equations at all points x, y, z at which $\mathbf{F}$, ϵ, and μ are continuous. For the purpose of determining the coefficients of the asymptotic expansions (6) we wish the $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ belonging to the pulse solution and this solution, as well as the other ones we are dealing with, is determined by the source.

In general there will be several (perhaps an infinite number of) functions

---

[20] See article 10 of Luneberg, Brown Notes, Volume I. It is also possible to obtain the wave-fronts from a knowledge of the rays. See articles 8 and 9 of this reference.

[21] *Cf.* p. 84 of Luneberg, Brown Notes, already referred to.

$\mathbf{A}_\nu(\tau)$, $\mathbf{B}_\nu(\tau)$ for a given ν. One pair, $\mathbf{A}_\nu$, $\mathbf{B}_\nu$, states the behavior of the discontinuities of the ν-th time derivatives of $\mathbf{E}$ and $\mathbf{H}$ along the rays belonging to the wave-fronts emanating from the source and for the moment we shall confine ourselves to this pair. The parameter τ in equations (46) can be time itself or something directly related to the time. At $\tau = 0$, the ray is at the source and so knowledge of $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ at $\tau = 0$ can be expected to depend upon knowledge of the source. Indeed it appears possible (compare footnotes 11 and 13) to determine $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ at $\tau = 0$ in terms of the source function $\mathbf{F}$ and its time derivatives. However, for the very important case of a dipole or point source the function $\mathbf{g}(x, y, z)$ in $\mathbf{F} = \mathbf{g}f(t)$ is the δ-function and it is awkward to work with it.

We shall therefore dispose of the problem of uniquely fixing the $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ of the field directly transmitted by the source in an alternative way which is consistent with the preceding remarks and which should give the same result. If these quantities were known at one point on a ray they could be fixed uniquely. Now the quantities $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ are to be substituted in the expressions for $\mathbf{u}$ and $\mathbf{v}$ given by (6). Hence we may look to the problem for which $\mathbf{u}$ and $\mathbf{v}$ give the solution to find suitable initial or boundary conditions for $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$.

If, for example, $\mathbf{u}$ and $\mathbf{v}$ represent the spatial factors of the steady state field of a dipole in a non-homogeneous medium, then we should want as a suitable boundary condition on $\mathbf{u}$ and $\mathbf{v}$ that they satisfy at the dipole the boundary conditions on the dipole when placed in a homogeneous medium. We could therefore require that each term of the asymptotic expansion for the $\mathbf{u}$ and $\mathbf{v}$ which represent the dipole field in a non-homogeneous medium have the same limiting value at the dipole as does the corresponding term of the asymptotic expansion for the dipole in a homogeneous medium. Now the asymptotic expansion for the dipole in a homogeneous medium can be obtained in other ways[22] and is

$$\mathbf{u} = \frac{\mathbf{m} \times \boldsymbol{\varrho}}{r} \exp\{ikr\} + \frac{(\mathbf{m} \times \boldsymbol{\varrho})i}{kr^2} \exp\{ikr\}$$

$$\mathbf{v} = \frac{\boldsymbol{\varrho} \times (\mathbf{m} \times \boldsymbol{\varrho})}{r} \exp\{ikr\} \tag{48}$$

$$+ \frac{(\boldsymbol{\varrho} \times (\mathbf{m} \times \boldsymbol{\varrho})i - (\mathbf{m}\cdot\boldsymbol{\varrho})\boldsymbol{\varrho}2i) \exp\{ikr\}}{kr^2} + \frac{(3(\mathbf{m}\cdot\boldsymbol{\varrho})\boldsymbol{\varrho} - \mathbf{m}) \exp\{ikr\}}{k^2r^3}$$

wherein the vector $\boldsymbol{\varrho} = (x, y, z)/r$ and $\mathbf{m}$ is a constant complex vector giving the moment and orientation of the dipole located at (0, 0, 0). The quantity r in $\exp\{ikr\}$ corresponds in a homogeneous medium, to the $\psi(x, y, z)$ of our theory.

Examination of equations (48) shows that as r approaches 0 each term becomes infinite. Hence the limiting values approached by each term of (48)

[22]Luneberg, Brown Notes, Volume I, p. 79.

cannot be used directly as boundary conditions; however, a simple reformulation can be. The expansion coefficient of rays in a non-homeogeneous medium[23], denoted by K, is equal to r^2 in a homogeneous medium. If we multiply the n-th term in each expansion of (48) by $(K^{1/2})^{n+1}$ (the first term given by $n = 0$), we see that each term approaches a finite limit. The asymptotic expansions of $\mathbf{u}$ and $\mathbf{v}$ for a dipole in a non-homogeneous medium will not contain simple inverse powers of r^2. These will be replaced by some quantities involving K. However, to meet the boundary condition suggested above we can require that the n-th term be multiplied by $(K^{1/2})^{n+1}$ and that the limit of this product as it approaches the dipole be the limit given by the corresponding term of (48) multiplied by r^{n+1}. But otherwise, the limit of $(K^{1/2})^{\nu+1}\mathbf{A}_\nu$ and $(K^{1/2})^{\nu+1}\mathbf{B}_\nu$ must approach the limits of the corresponding terms of (48) multiplied by $r^{\nu+1}$. It will be noted that except for the first two values of ν in the case of $\mathbf{u}$ and the first three in the case of $\mathbf{v}$, the limits which $(K^{1/2})^{\nu+1}\mathbf{A}_\nu$ and $(K^{1/2})^{\nu+1}\mathbf{B}_\nu$ must satisfy are zero.

This method of determining the arbitrary constants in the expressions for $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ takes care of that summand of the ν-th coefficient of $\mathbf{u}$ and $\mathbf{v}$ of (6) which corresponds to the ψ_α of the directly transmitted set of wave-fronts. For wave-fronts which arise from other causes, such as the presence of discontinuities in ϵ and μ, other initial conditions now being investigated must be used.

7. *Conclusion*

Theoretically the problem of obtaining the successive coefficients of the asymptotic expansions for $\mathbf{u}$ and $\mathbf{v}$, the spatial factors of time harmonic $\mathbf{E}$ and $\mathbf{H}$, can be solved without knowing the full pulse solution $\mathbf{E}_0$, $\mathbf{H}_0$. By utilizing the fact that the coefficients depend only on the discontinuities of $\mathbf{E}_0$, $\mathbf{H}_0$ and their time derivatives on the wave-fronts passing any point in space, it is possible to obtain a system of first order ordinary differential equations whose solutions are these coefficients. Solution of even the first two sets of equations of this recursive system, that is the cases $\nu = 0$ and $\nu = 1$, would give a significant improvement over geometrical optics approximations to time harmonic fields.

The theory of this paper can be applied at least theoretically to find the asymptotic form of the dipole field in a non-homogeneous medium. There is little doubt that it can be carried further to treat problems in which more general media and boundary conditions beyond that of the dipole are involved. However, further work is needed on this extension of the theory. A difficulty which at the present writing is not resolved is that one must know all the families of wave fronts and rays which arise from the source and boundaries, for contributions to each coefficient of the asymptotic expansion will come from each wave-front through a point in space.

---

[23]For the definition of K see Luneberg, Brown Notes, Volume I, p. 50. K is the reciprocal of the Gaussian curvature of the wave-front.

Appendix

Derivation of the Ordinary Differential Equations for $\mathbf{A}_\nu$ *and* $\mathbf{B}_\nu$

This appendix will derive the ordinary differential equations (46) for $\mathbf{A}_\nu$ and $\mathbf{B}_\nu$ by utilizing equations (41) and (45). Equations (45) are restated as

$$\mathbf{A}_0 \cdot \operatorname{curl} \mathbf{B}_\nu - \mathbf{B}_0 \cdot \operatorname{curl} \mathbf{A}_\nu = 0, \tag{1}$$

$$\epsilon \mathbf{A}_0 \cdot \operatorname{curl} \mathbf{A}_\nu + \mu \mathbf{B}_0 \cdot \operatorname{curl} \mathbf{B}_\nu = 0. \tag{2}$$

Equations (41) are

$$\operatorname{grad} \psi \times \mathbf{B}_\nu + \epsilon \mathbf{A}_\nu = c \operatorname{curl} \mathbf{B}_{\nu-1} , \tag{3}$$

$$\operatorname{grad} \psi \times \mathbf{A}_\nu - \mu \mathbf{B}_\nu = c \operatorname{curl} \mathbf{A}_{\nu-1} . \tag{4}$$

Dot both sides of equations (3) and (4) by grad ψ. Then

$$\epsilon \operatorname{grad} \psi \cdot \mathbf{A}_\nu = c \operatorname{grad} \psi \cdot \operatorname{curl} \mathbf{B}_{\nu-1} , \tag{5}$$

$$\mu \operatorname{grad} \psi \cdot \mathbf{B}_\nu - -c \operatorname{grad} \psi \cdot \operatorname{curl} \mathbf{A}_{\nu-1} . \tag{6}$$

Taking the divergence[24] of both sides of (3) and (4), and using the fact that div curl $= 0$, we get

$$\operatorname{div} \epsilon \mathbf{A}_\nu = -\operatorname{div} (\operatorname{grad} \psi \times \mathbf{B}_\nu) = +\operatorname{grad} \psi \cdot \operatorname{curl} \mathbf{B}_\nu , \tag{7}$$

$$\operatorname{div} \mu \mathbf{B}_\nu = \operatorname{div} (\operatorname{grad} \psi \times \mathbf{A}_\nu) = -\operatorname{grad} \psi \cdot \operatorname{curl} \mathbf{A}_\nu . \tag{8}$$

From (5) and (6),

$$\operatorname{grad} \psi \cdot \mathbf{A}_\nu = \frac{c}{\epsilon} \operatorname{grad} \psi \cdot \operatorname{curl} \mathbf{B}_{\nu-1} , \tag{9}$$

$$\operatorname{grad} \psi \cdot \mathbf{B}_\nu = -\frac{c}{\mu} \operatorname{grad} \psi \cdot \operatorname{curl} \mathbf{A}_{\nu-1} . \tag{10}$$

Using the first and third members of (7) and (8) for $\nu = \nu - 1$ and substituting in (9) and (10), respectively, gives

---

[24]In taking the divergence of both sides of (3) and (4) we must recall that the curl operator in (3) and (4) applies to $\mathbf{A}_{\nu-1}$ and $\mathbf{B}_{\nu-1}$ as functions of x, y, z and $\psi(x, y, z)$. Hence the divergence is to apply in the same sense to the functions on the left. We therefore understand hereafter that in all quantities $\mathbf{A}_\nu$, $\mathbf{B}_\nu$ for all ν, t is first replaced by $\psi(x, y, z)/c$. This is the correct sense of equations (1) to (4) for they hold on the discontinuity surface $\psi(x, y, z) - ct = 0$

$$\text{grad}\ \psi \cdot \mathbf{A}_\nu = \frac{c}{\epsilon}\ \text{div}\ \epsilon \mathbf{A}_{\nu-1}\,, \tag{11}$$

$$\text{grad}\ \psi \cdot \mathbf{B}_\nu = \frac{c}{\mu}\ \text{div}\ \mu \mathbf{B}_{\nu-1}\,. \tag{12}$$

From (4),

$$\mathbf{B}_\nu = \frac{\text{grad}\ \psi}{\mu} \times \mathbf{A}_\nu - \frac{c}{\mu}\ \text{curl}\ \mathbf{A}_{\nu-1}\,. \tag{13}$$

We substitute this in (2) and obtain

$$\epsilon \mathbf{A}_0 \cdot \text{curl}\ \mathbf{A}_\nu + \mu \mathbf{B}_0 \cdot \text{curl} \left(\frac{\text{grad}\ \psi}{\mu} \times \mathbf{A}_\nu \right) - \mu \mathbf{B}_0 \cdot \text{curl} \left(\frac{c}{\mu}\ \text{curl}\ \mathbf{A}_{\nu-1} \right) = 0. \tag{14}$$

Now $\epsilon \mathbf{A}_0 = -\text{grad}\ \psi \times \mathbf{B}_0$ by equation (23) of the text proper. Hence

$$\begin{aligned} &\mathbf{B}_0 \times \text{grad}\ \psi \cdot \text{curl}\ \mathbf{A}_\nu + \mu \mathbf{B}_0 \cdot \text{curl} \left(\frac{\text{grad}\ \psi}{\mu} \times \mathbf{A}_\nu \right) \\ &\qquad - \mu \mathbf{B}_0 \cdot \text{curl} \left(\frac{c}{\mu}\ \text{curl}\ \mathbf{A}_{\nu-1} \right) = 0. \end{aligned} \tag{15}$$

Dividing by μ and interchanging dot and cross give

$$\begin{aligned} &\mathbf{B}_0 \cdot \frac{\text{grad}\ \psi}{\mu} \times \text{curl}\ \mathbf{A}_\nu + \mathbf{B}_0 \cdot \text{curl} \left(\frac{\text{grad}\ \psi}{\mu} \times \mathbf{A}_\nu \right) \\ &\qquad - \mathbf{B}_0 \cdot \text{curl} \left(\frac{c}{\mu}\ \text{curl}\ \mathbf{A}_{\nu-1} \right) = 0. \end{aligned} \tag{16}$$

Now by a vector identity and the fact that curl grad $= 0$,

$$\mathbf{A}_\nu \times \text{curl} \left(\frac{\text{grad}\ \psi}{\mu} \right) = \mathbf{A}_\nu \times \left(\text{grad}\ \frac{1}{\mu} \times \text{grad}\ \psi \right).$$

Dotting both sides by $\mathbf{B}_0$ gives

$$\mathbf{B}_0 \cdot \left(\mathbf{A}_\nu \times \text{curl}\ \frac{\text{grad}\ \psi}{\mu} \right) = \mathbf{B}_0 \cdot \left(\mathbf{A}_\nu \times \left(\text{grad}\ \frac{1}{\mu} \times \text{grad}\ \psi \right) \right). \tag{17}$$

Now let us add (17) to both sides of (16) and transfer the last term on the left side of (16) to the right side. We obtain, after rearranging terms,

$$\begin{aligned} &\mathbf{B}_0 \cdot \left\{ \text{curl} \left(\frac{\text{grad}\ \psi}{\mu} \times \mathbf{A}_\nu \right) + \frac{\text{grad}\ \psi}{\mu} \times \text{curl}\ \mathbf{A}_\nu + \mathbf{A}_\nu \times \text{curl}\ \frac{\text{grad}\ \psi}{\mu} \right\} \\ &\qquad = \mathbf{B}_0 \cdot \text{curl} \left(\frac{c}{\mu}\ \text{curl}\ \mathbf{A}_{\nu-1} \right) + \mathbf{B}_0 \cdot \left\{ \mathbf{A}_\nu \times \left(\text{grad}\ \frac{1}{\mu} \times \text{grad}\ \psi \right) \right\}. \end{aligned} \tag{18}$$

We shall now use the following vector identity,[25]

$$\operatorname{curl}(\mathbf{A} \times \mathbf{B}) + \mathbf{A} \times \operatorname{curl} \mathbf{B} + B \times \operatorname{curl} \mathbf{A}$$

$$= -2 \frac{\partial \mathbf{B}}{\partial \alpha} - \mathbf{B} \operatorname{div} \mathbf{A} + \mathbf{A} \operatorname{div} \mathbf{B} + \operatorname{grad} (\mathbf{A} \cdot \mathbf{B}).$$

In this identity

$$\frac{\partial}{\partial \alpha} = A_1 \frac{\partial}{\partial x} + A_2 \frac{\partial}{\partial y} + A_3 \frac{\partial}{\partial z} = (\mathbf{A} \cdot \nabla)$$

where A_1, A_2, and A_3 are the components of $\mathbf{A}$; the symbol $\partial/\partial\alpha$ denotes differentiation in the direction of the vector $\mathbf{A}$. In our case $\mathbf{A} = (1/\mu) \operatorname{grad} \psi$ and $\mathbf{B} = \mathbf{A}_\nu$. Then

$$A_1 = \frac{1}{\mu} \psi_x, \quad A_2 = \frac{1}{\mu} \psi_y, \quad A_3 = \frac{1}{\mu} \psi_z. \qquad \text{Let } \frac{\partial}{\partial \tau} = \psi_x \frac{\partial}{\partial x} + \psi_y \frac{\partial}{\partial y} + \psi_z \frac{\partial}{\partial z}.$$

Then from (18)

$$\mathbf{B}_0 \cdot \left\{ -\frac{2}{\mu} \frac{\partial \mathbf{A}_\nu}{\partial \tau} - \mathbf{A}_\nu \operatorname{div} \left(\frac{\operatorname{grad} \psi}{\mu} \right) + \frac{\operatorname{grad} \psi}{\mu} \operatorname{div} \mathbf{A}_\nu + \operatorname{grad} \left(\frac{\operatorname{grad} \psi}{\mu} \cdot \mathbf{A}_\nu \right) \right\}$$

$$= \mathbf{B}_0 \cdot \operatorname{curl} \left(\frac{c}{\mu} \operatorname{curl} \mathbf{A}_{\nu-1} \right) + \mathbf{B}_0 \cdot \left\{ \mathbf{A}_\nu \times \left(\operatorname{grad} \frac{1}{\mu} \times \operatorname{grad} \psi \right) \right\},$$

(19)

$$= \mathbf{B}_0 \cdot \operatorname{curl} \left(\frac{c}{\mu} \operatorname{curl} \mathbf{A}_{\nu-1} \right)$$

$$+ \mathbf{B}_0 \cdot \left\{ (\mathbf{A}_\nu \cdot \operatorname{grad} \psi) \operatorname{grad} \frac{1}{\mu} - \left(\mathbf{A}_\nu \cdot \operatorname{grad} \frac{1}{\mu} \right) \operatorname{grad} \psi \right\}.$$

By a standard vector identity

$$\operatorname{grad} \left(\frac{1}{\mu} (\operatorname{grad} \psi \cdot \mathbf{A}_\nu) \right) = \operatorname{grad} \frac{1}{\mu} (\operatorname{grad} \psi \cdot \mathbf{A}_\nu) + \frac{1}{\mu} \operatorname{grad} (\operatorname{grad} \psi \cdot \mathbf{A}_\nu).$$

---

[25]This identity results from adding the standard vector identity for curl $(\mathbf{A} \times \mathbf{B})$ and an identity for $\mathbf{A} \times \operatorname{curl} \mathbf{B} + \mathbf{B} \times \operatorname{curl} \mathbf{A}$. The latter identities are readily proved.

$$\mathbf{A} \times \operatorname{curl} \mathbf{B} = \mathbf{A} \times \left(i \times \frac{\partial \mathbf{B}}{\partial x} \right) + \mathbf{A} \times \left(j \times \frac{\partial \mathbf{B}}{\partial y} \right) + \mathbf{A} \times \left(k \times \frac{\partial \mathbf{B}}{\partial z} \right)$$

$$= \left(\mathbf{A} \cdot \frac{\partial \mathbf{B}}{\partial x} \right) i + \left(\mathbf{A} \cdot \frac{\partial \mathbf{B}}{\partial y} \right) j + \left(\mathbf{A} \cdot \frac{\partial \mathbf{B}}{\partial z} \right) - A_1 \frac{\partial \mathbf{B}}{\partial x} - A_2 \frac{\partial \mathbf{B}}{\partial y} - A_3 \frac{\partial \mathbf{B}}{\partial z}.$$

Likewise we form $\mathbf{B} \times \operatorname{curl} \mathbf{A}$ and add. See Appendix I, p. 7, of Luneberg, Brown Notes, Volume II.

If we replace the last term on the left side of (19) by the equivalent just indicated we note that a term on the new left side cancels one on the right. Also since $\mathbf{B}_0 \cdot \operatorname{grad} \psi = 0$ we may throw out some terms and obtain

$$(20) \quad \begin{aligned} &\mathbf{B}_0 \cdot \left\{ -\frac{2}{\mu} \frac{\partial \mathbf{A}_\nu}{\partial \tau} - \mathbf{A}_\nu \operatorname{div} \left(\frac{\operatorname{grad} \psi}{\mu} \right) + \frac{1}{\mu} \operatorname{grad} (\operatorname{grad} \psi \cdot \mathbf{A}_\nu) \right\} \\ &= B_0 \cdot \operatorname{curl} \left(\frac{c}{\mu} \operatorname{curl} \mathbf{A}_{\nu-1} \right). \end{aligned}$$

By (11) we have

$$(21) \quad \begin{aligned} &\mathbf{B}_0 \cdot \left\{ -\frac{2}{\mu} \frac{\partial \mathbf{A}_\nu}{\partial \tau} - \mathbf{A}_\nu \operatorname{div} \left(\frac{\operatorname{grad} \psi}{\mu} \right) \right\} \\ &= +\mathbf{B}_0 \cdot \left\{ \operatorname{curl} \left(\frac{c}{\mu} \operatorname{curl} \mathbf{A}_{\nu-1} \right) - \frac{1}{\mu} \operatorname{grad} \left(\frac{c}{\epsilon} \operatorname{div} \epsilon \mathbf{A}_{\nu-1} \right) \right\}. \end{aligned}$$

Now

$$\operatorname{div} \left(\frac{\operatorname{grad} \psi}{\mu} \right) = \frac{\partial}{\partial x} \left(\frac{\psi_x}{\mu} \right) + \cdots = \frac{1}{\mu} \Delta_\mu \psi$$

by definition of the symbol $\Delta_\mu \psi$. Hence, multiplying by μ and a minus sign in (21) gives

$$(22) \quad \mathbf{B}_0 \cdot \left(2 \frac{\partial \mathbf{A}_\nu}{\partial \tau} \right) + (\mathbf{A}_\nu \Delta_\mu \psi) = -\mathbf{B}_0 \cdot \left\{ \mu \operatorname{curl} \left(\frac{c}{\mu} \operatorname{curl} \mathbf{A}_{\nu-1} \right) - \operatorname{grad} \left(\frac{c}{\epsilon} \operatorname{div} \epsilon \mathbf{A}_{\nu-1} \right) \right\}.$$

We may rewrite equation (22) as

$$(22a) \quad \mathbf{B}_0 \cdot \left\{ 2 \frac{\partial \mathbf{A}_\nu}{\partial \tau} + \mathbf{A}_\nu \Delta_\mu \psi + \mu \operatorname{curl} \left(\frac{c}{\mu} \operatorname{curl} \mathbf{A}_{\nu-1} \right) - \operatorname{grad} \left(\frac{c}{\epsilon} \operatorname{div} \epsilon \mathbf{A}_{\nu-1} \right) \right\} = 0.$$

In so far as $\partial \mathbf{A}_\nu / \partial \tau$ is concerned, this is a derivative in the direction of grad ψ. If we regard x, y, z in $\mathbf{A}_\nu$ as functions of τ, the parameter along the ray perpendicular to grad ψ, we may write instead $d\mathbf{A}_\nu / d\tau$.

We note that the second factor in equation (22a) is a vector perpendicular to $\mathbf{B}_0$ and hence in the plane of $\mathbf{A}_0$ and grad ψ. We can go through a series of steps analogous to those above to show that this quantity is also perpendicular to $\mathbf{A}_0$ and so must have the direction of grad ψ. The steps are readily indicated, for the changes over what was done above are minor. We need go back only to step (13) which gives an expression for $\mathbf{B}_\nu$ and substitute this value of $\mathbf{B}_\nu$ in (1). Then

$$(23) \quad \mathbf{A}_0 \cdot \left\{ \operatorname{curl} \left(\frac{\operatorname{grad} \psi}{\mu} \times \mathbf{A}_\nu \right) - \operatorname{curl} \left(\frac{c}{\mu} \operatorname{curl} \mathbf{A}_{\nu-1} \right) \right\} - \mathbf{B}_0 \cdot \operatorname{curl} \mathbf{A}_\nu = 0.$$

From equation (24) of the text proper $\mathbf{B}_0 = (1/\mu) \text{ grad } \psi \times \mathbf{A}_0 = -(1/\mu)\mathbf{A}_0 \times \text{grad } \psi$. Hence

$$(24) \quad \mathbf{A}_0 \cdot \left\{ \text{curl}\left(\frac{\text{grad } \psi}{\mu} \times \mathbf{A}_\nu\right) - \text{curl}\left(\frac{c}{\mu} \text{ curl } \mathbf{A}_{\nu-1}\right)\right\} + \mathbf{A}_0 \times \frac{\text{grad } \psi}{\mu} \cdot \text{curl } \mathbf{A}_\nu = 0.$$

Interchange dot and cross in last term. If we compare (24) and (16) we see that we can repeat everything we did from (16) on to (22) and conclude that (22a) holds with $\mathbf{A}_0$ replacing $\mathbf{B}_0$. Hence the second factor in equation (22a) is a vector perpendicular to both $\mathbf{A}_0$ and $\mathbf{B}_0$. Since $\mathbf{A}_0$, $\mathbf{B}_0$, and grad ψ are mutually perpendicular, this factor is parallel to grad ψ. We may say therefore that

$$(25) \quad 2\frac{d\mathbf{A}_\nu}{d\tau} + \mathbf{A}_\nu \Delta_\mu \psi + \mu \text{ curl}\left(\frac{c}{\mu} \text{ curl } \mathbf{A}_{\nu-1}\right) - \text{grad}\left(\frac{c}{\epsilon} \text{ div } \epsilon\mathbf{A}_{\nu-1}\right) = R \text{ grad } \psi,$$

where R is a scalar function to be determined. To save writing we let $\mathbf{C}_\nu = \mu \text{ curl } (c/\mu \text{ curl } \mathbf{A}_{\nu-1}) - \text{grad } (c/\epsilon \text{ div } \epsilon\mathbf{A}_{\nu-1})$. Then

$$(26) \qquad 2\frac{d\mathbf{A}_\nu}{d\tau} + \mathbf{A}_\nu \Delta_\mu \psi + \mathbf{C}_\nu = R \text{ grad } \psi.$$

$\mathbf{C}_\nu$ thus stands for two terms, which will not be affected by what follows.

We now determine the value of R. Dot both sides of (26) by grad ψ. This gives

$$(27) \qquad 2\frac{d\mathbf{A}_\nu}{d\tau} \cdot \text{grad } \psi + \Delta_\mu \psi \mathbf{A}_\nu \cdot \text{grad } \psi + \mathbf{C}_\nu \cdot \text{grad } \psi = Rn^2.$$

Using the fact that $n = (\psi_x^2 + \psi_y^2 + \psi_z^2)^{1/2}$ we prove readily, since

$$\frac{d}{d\tau} = \psi_x \frac{\partial}{\partial x} + \psi_y \frac{\partial}{\partial y} + \psi_z \frac{\partial}{\partial z},$$

that

$$(28) \qquad n \text{ grad } n = \frac{d}{d\tau}(\text{grad } \psi).$$

Hence

$$(29) \qquad \frac{d}{d\tau}(\mathbf{A}_\nu \cdot \text{grad } \psi) = \frac{d\mathbf{A}_\nu}{d\tau} \cdot \text{grad } \psi + \mathbf{A}_\nu \cdot n \text{ grad } n.$$

We may therefore rewrite the value of Rn^2 in (27) as

$$(30) \quad Rn^2 = -2n\mathbf{A}_\nu \cdot \text{grad } n + 2\frac{d}{d\tau}(\mathbf{A}_\nu \cdot \text{grad } \psi) + \Delta_\mu \psi \mathbf{A}_\nu \cdot \text{grad } \psi + \mathbf{C}_\nu \cdot \text{grad } \psi.$$

Hence

$$R = -\frac{2}{n}\operatorname{grad} n\cdot\mathbf{A}_\nu + P_\nu \tag{31}$$

where

$$P_\nu = \frac{2 d}{n^2 d\tau}(\mathbf{A}_\nu\cdot\operatorname{grad}\psi) + \Delta_\mu\psi\frac{\mathbf{A}_\nu\cdot\operatorname{grad}\psi}{n^2} + \frac{\mathbf{C}_\nu\cdot\operatorname{grad}\psi}{n^2}. \tag{32}$$

We shall show that $P_\nu = 0$. From (3), wherein we dot both sides by grad ψ, we obtain

$$\epsilon\mathbf{A}_\nu\cdot\operatorname{grad}\psi = c\operatorname{curl}\mathbf{B}_{\nu-1}\cdot\operatorname{grad}\psi. \tag{33}$$

Using equation (11) gives

$$\operatorname{div}\epsilon\mathbf{A}_{\nu-1} = \operatorname{curl}\mathbf{B}_{\nu-1}\cdot\operatorname{grad}\psi. \tag{34}$$

Since our relations hold for all $\nu \geq 0$,

$$\operatorname{div}\epsilon\mathbf{A}_\nu = \operatorname{curl}\mathbf{B}_\nu\cdot\operatorname{grad}\psi. \tag{35}$$

If we substitute the value of $\mathbf{B}_\nu$ from (4) into (35) we get

$$\operatorname{div}\epsilon\mathbf{A}_\nu = \operatorname{curl}\left(\frac{\operatorname{grad}\psi}{\mu}\times\mathbf{A}_\nu\right)\cdot\operatorname{grad}\psi - \operatorname{curl}\left(\frac{c}{\mu}\operatorname{curl}\mathbf{A}_{\nu-1}\right)\cdot\operatorname{grad}\psi. \tag{36}$$

Now let

$$\begin{aligned}\left[\frac{\operatorname{grad}\psi}{\mu}, \mathbf{A}_\nu\right] &= \operatorname{curl}\left(\frac{\operatorname{grad}\psi}{\mu}\times\mathbf{A}_\nu\right)\\ &\quad + \frac{\operatorname{grad}\psi}{\mu}\times\operatorname{curl}\mathbf{A}_\nu + \mathbf{A}_\nu\times\operatorname{curl}\frac{\operatorname{grad}\psi}{\mu}.\end{aligned} \tag{37}$$

Dot both sides of (37) by grad ψ and use (36). Then

$$\begin{aligned}\operatorname{div}\epsilon\mathbf{A}_\nu &= \left[\frac{\operatorname{grad}\psi}{\mu}, \mathbf{A}_\nu\right]\cdot\operatorname{grad}\psi - \left(\mathbf{A}_\nu\times\operatorname{curl}\frac{\operatorname{grad}\psi}{\mu}\right)\cdot\operatorname{grad}\psi\\ &\quad - \operatorname{curl}\left(\frac{c}{\mu}\operatorname{curl}\mathbf{A}_{\nu-1}\right)\cdot\operatorname{grad}\psi.\end{aligned} \tag{38}$$

Using a vector identity for the curl of a scalar and a vector in the second term on the right side gives

$$\begin{aligned}\operatorname{did}\epsilon\mathbf{A}_\nu &= \left[\frac{\operatorname{grad}\psi}{\mu}, \mathbf{A}_\tau\right]\cdot\operatorname{grad}\psi - \left(\mathbf{A}_\nu\times\left(\operatorname{grad}\frac{1}{\mu}\times\operatorname{grad}\psi\right)\right)\cdot\operatorname{grad}\psi\\ &\quad - \operatorname{curl}\left(\frac{c}{\mu}\operatorname{curl}\mathbf{A}_{\nu-1}\right)\cdot\operatorname{grad}\psi.\end{aligned} \tag{39}$$

Apply the vector identity for the vector triple product to the second term on the right side. We obtain

$$\text{div}\, \epsilon \mathbf{A}_\nu = \left[\frac{\text{grad}\, \psi}{\mu}, \mathbf{A}_\nu\right]\cdot \text{grad}\, \psi - (\mathbf{A}_\nu \cdot \text{grad}\, \psi)\left(\text{grad}\, \frac{1}{\mu}\cdot \text{grad}\, \psi\right) \tag{40}$$
$$+ \left(\mathbf{A}_\nu \cdot \text{grad}\, \frac{1}{\mu}\right)n^2 - \text{curl}\left(\frac{c}{\mu}\,\text{curl}\, \mathbf{A}_{\nu-1}\right)\cdot \text{grad}\, \psi.$$

Now the quantity $[(\text{grad}\, \psi)/\mu,\ \mathbf{A}_\nu]$ given by (37) is precisely the quantity encountered in (18) and converted by a vector identity into the form in (19). If we make the same change in (40), expand grad $[(\text{grad}\, \psi)/\mu \cdot \mathbf{A}_\nu]$ into grad $1/\mu$ $(\text{grad}\, \psi \cdot \mathbf{A}_\nu) + (1/\mu)$ grad $(\text{grad}\, \psi \cdot \mathbf{A}_\nu)$, and cancel terms we get

$$\text{div}\, \epsilon \mathbf{A}_\nu = -\frac{2}{\mu}\frac{d\mathbf{A}_\nu}{d\tau}\cdot \text{grad}\, \psi - (\mathbf{A}_\nu \cdot \text{grad}\, \psi)\, \text{div}\left(\frac{\text{grad}\, \psi}{\mu}\right)$$
$$+ \epsilon\, \text{div}\, \mathbf{A}_\nu + \frac{1}{\mu}\, \text{grad}\, (\mathbf{A}_\nu \cdot \text{grad}\, \psi)\cdot \text{grad}\, \psi \tag{41}$$
$$+ n^2\left(\mathbf{A}_\nu \cdot \text{grad}\, \frac{1}{\mu}\right) - \text{curl}\left(\frac{c}{\mu}\,\text{curl}\, \mathbf{A}_{\nu-1}\right)\cdot \text{grad}\, \psi.$$

By a vector identity, $\text{div}\, \epsilon \mathbf{A}_\nu = \mathbf{A}_\nu \cdot \text{grad}\, \epsilon + \epsilon\, \text{div}\, \mathbf{A}_\nu$. If we use this equation in (41) and multiply by μ, we obtain

$$0 = -2\,\frac{d\mathbf{A}_\nu}{d\tau}\cdot \text{grad}\, \psi - \mu(\mathbf{A}_\nu \cdot \text{grad}\, \psi)\, \text{div}\left(\frac{\text{grad}\, \psi}{\mu}\right)$$
$$+ \text{grad}\, (\mathbf{A}_\nu \cdot \text{grad}\, \psi)\cdot \text{grad}\, \psi - \mu \mathbf{A}_\nu \cdot \text{grad}\, \epsilon \tag{42}$$
$$+ \mu n^2\left(\mathbf{A}_\nu \cdot \text{grad}\, \frac{1}{\mu}\right) - \mu\, \text{curl}\left(\frac{c}{\mu}\,\text{curl}\, \mathbf{A}_{\nu-1}\right)\cdot \text{grad}\, \psi.$$

We now use the fact noted above in connection with (22), namely, that $\mu\, \text{div}\, [(\text{grad}\, \psi)/\mu)] = \Delta_\mu \psi$. Also, since

$$\frac{d}{d\tau} = \psi_x \cdot \frac{\partial}{\partial x} + \psi_y\, \frac{\partial}{\partial y} + \psi_z\, \frac{\partial}{\partial z},$$

the quantity grad $f \cdot$ grad ψ, where f is any scalar, is $df/d\tau$. Hence

$$\text{grad}\, (\mathbf{A}_\nu \cdot \text{grad}\, \psi)\cdot \text{grad}\, \psi = \frac{d}{d\tau}\, (\mathbf{A}_\nu \cdot \text{grad}\, \psi). \tag{43}$$

The use of these several facts in (42) gives

$$
\begin{aligned}
0 = &-2\frac{d\mathbf{A}_\nu}{d\tau}\cdot \operatorname{grad}\psi - \Delta_\mu\psi(\mathbf{A}_\nu\cdot\operatorname{grad}\psi) \\
&+\frac{d}{d\tau}(\mathbf{A}_\nu\cdot\operatorname{grad}\psi) - \mu\mathbf{A}_\nu\cdot\operatorname{grad}\epsilon \\
&+\mu n^2\left(\mathbf{A}_\nu\cdot\operatorname{grad}\frac{1}{\mu}\right) - \mu\operatorname{curl}\left(\frac{c}{\mu}\operatorname{curl}\mathbf{A}_{\nu-1}\right)\cdot\operatorname{grad}\psi.
\end{aligned}
\tag{44}
$$

If we use (29) to eliminate the first term in (44) we obtain

$$
\begin{aligned}
0 = &-\frac{d}{d\tau}(\mathbf{A}_\nu\cdot\operatorname{grad}\psi) + 2\mathbf{A}_\nu\cdot n\operatorname{grad}n \\
&-\Delta_\mu\psi(\mathbf{A}_\nu\cdot\operatorname{grad}\psi) - \mu\mathbf{A}_\nu\cdot\operatorname{grad}\epsilon \\
&-\epsilon(\mathbf{A}_\nu\cdot\operatorname{grad}\mu) - \mu\operatorname{curl}\left(\frac{c}{\mu}\operatorname{curl}\mathbf{A}_{\nu-1}\right)\cdot\operatorname{grad}\psi.
\end{aligned}
\tag{45}
$$

Because $2n\operatorname{grad}n = \operatorname{grad}n^2 = \operatorname{grad}\epsilon\mu = \epsilon\operatorname{grad}\mu + \mu\operatorname{grad}\epsilon$, the second, fourth and fifth terms cancel and we have

$$
0 = +\frac{d}{d\tau}(\mathbf{A}_\nu\cdot\operatorname{grad}\psi) + \Delta_\mu\psi(\mathbf{A}_\nu\cdot\operatorname{grad}\psi) + \mu\operatorname{curl}\left(\frac{c}{\mu}\operatorname{curl}\mathbf{A}_{\nu-1}\right)\cdot\operatorname{grad}\psi.
\tag{46}
$$

From (32) and the value of $\mathbf{C}_\nu$ it follows that

$$
\begin{aligned}
n^2P_\nu = &\,2\frac{d}{d\tau}(\mathbf{A}_\nu\cdot\operatorname{grad}\psi) + \Delta_\mu\psi(\mathbf{A}_\nu\cdot\operatorname{grad}\psi) + \mu\operatorname{curl}\left(\frac{c}{\mu}\operatorname{curl}\mathbf{A}_{\nu-1}\right)\cdot\operatorname{grad}\psi \\
&-\operatorname{grad}\left(\frac{c}{\epsilon}\operatorname{div}(\epsilon\mathbf{A}_{\nu-1})\right)\cdot\operatorname{grad}\psi.
\end{aligned}
\tag{47}
$$

By (11) and (43) we may replace the last term on the right by $(\mathbf{A}_\nu\cdot\operatorname{grad}\psi)\,d/d\tau$. Hence, comparing (46) and (47), we see that P_ν is zero.

We note this fact in (31) and then substitute the value of R from (31) in (26). We have finally

$$
2\frac{d\mathbf{A}_\nu}{d\tau} + \mathbf{A}_\nu\Delta_\mu\psi + \frac{2}{n}(\operatorname{grad}n\cdot\mathbf{A}_\nu)\operatorname{grad}\psi = -\mathbf{C}_\nu
\tag{48}
$$

where $\mathbf{C}_\nu = \mu\operatorname{curl}((c/\mu)\operatorname{curl}\mathbf{A}_{\nu-1}) - \operatorname{grad}((c/\epsilon)\operatorname{div}\epsilon\mathbf{A}_{\nu-1})$.

To obtain the analogous ordinary differential equation for $\mathbf{B}_\nu$ we note first that steps (1) to (12) treat both quantities and are basic to (48) as well as the equation we now seek. Starting then with a step analogous to (13), this time we substitute the value of $\mathbf{A}_\nu$ from (3) into (1) and obtain

$$
\mathbf{A}_0\cdot\operatorname{curl}\mathbf{B}_\nu - \mathbf{B}_0\cdot\left\{\operatorname{curl}\left(-\frac{1}{\epsilon}\operatorname{grad}\psi\times\mathbf{B}_\nu\right) + \operatorname{curl}\left(\frac{c}{\epsilon}\operatorname{curl}\mathbf{B}_{\nu-1}\right)\right\} = 0.
\tag{14'}
$$

Again we replace $\mathbf{A}_0$ by $-1/\epsilon \operatorname{grad} \psi \times \mathbf{B}_0$ and obtain

$$
(15') \quad \begin{aligned}
&-\frac{1}{\epsilon} \operatorname{grad} \psi \times \mathbf{B}_0 \cdot \operatorname{curl} \mathbf{B}_\nu \\
&- \mathbf{B}_0 \cdot \left\{ \operatorname{curl} \left(-\frac{1}{\epsilon} \operatorname{grad} \psi \times \mathbf{B}_\nu \right) + \operatorname{curl} \left(\frac{c}{\epsilon} \operatorname{curl} \mathbf{B}_{\nu-1} \right) \right\} = 0.
\end{aligned}
$$

$$
(16') \quad \begin{aligned}
&\mathbf{B}_0 \cdot \frac{\operatorname{grad} \psi}{\epsilon} \times \operatorname{curl} \mathbf{B}_\nu + \\
&\mathbf{B}_0 \cdot \operatorname{curl} \left(\frac{\operatorname{grad} \psi}{\epsilon} \times \mathbf{B}_\nu \right) - \mathbf{B}_0 \cdot \operatorname{curl} \left(\frac{c}{\epsilon} \operatorname{curl} \mathbf{B}_{\nu-1} \right) = 0.
\end{aligned}
$$

Analogously to the derivation of (17) we derive

$$
(17') \quad \mathbf{B}_0 \cdot \left(\mathbf{B}_\nu \times \operatorname{curl} \frac{\operatorname{grad} \psi}{\epsilon} \right) = \mathbf{B}_0 \cdot \left\{ \mathbf{B}_\nu \times \left(\operatorname{grad} \frac{1}{\epsilon} \times \operatorname{grad} \psi \right) \right\}.
$$

We add both sides of (17′) to (16′), transfer the last term of (16′) to the right side and obtain

$$
(18') \quad \begin{aligned}
&\mathbf{B}_0 \cdot \left\{ \operatorname{curl} \left(\frac{\operatorname{grad} \psi}{\epsilon} \times \mathbf{B}_\nu \right) + \frac{\operatorname{grad} \psi}{\epsilon} \times \operatorname{curl} \mathbf{B}_\nu + \mathbf{B}_\nu \times \operatorname{curl} \frac{\operatorname{grad} \psi}{\epsilon} \right\} \\
&\quad = \mathbf{B}_0 \cdot \operatorname{curl} \left(\frac{c}{\epsilon} \operatorname{curl} \mathbf{B}_{\nu-1} \right) + \mathbf{B}_0 \cdot \left\{ \mathbf{B}_\nu \times \left(\operatorname{grad} \frac{1}{\epsilon} \times \operatorname{grad} \psi \right) \right\}.
\end{aligned}
$$

Again we use the vector identity as in the step from (18) to (19) and obtain

$$
(19') \quad \begin{aligned}
&\mathbf{B}_0 \cdot \left\{ -\frac{2}{\epsilon} \frac{\partial \mathbf{B}_\nu}{\partial \tau} - \mathbf{B}_\nu \operatorname{div} \left(\frac{\operatorname{grad} \psi}{\epsilon} \right) + \frac{\operatorname{grad} \psi}{\epsilon} \operatorname{div} \mathbf{B}_\nu + \operatorname{grad} \left(\frac{\operatorname{grad} \psi}{\epsilon} \cdot \mathbf{B}_\nu \right) \right\} \\
&\quad = \mathbf{B}_0 \cdot \operatorname{curl} \left(\frac{c}{\epsilon} \operatorname{curl} \mathbf{B}_{\nu-1} \right) + \mathbf{B}_0 \cdot \left\{ (\mathbf{B}_\nu \times \left(\operatorname{grad} \frac{1}{\epsilon} \times \operatorname{grad} \psi \right) \right\}, \\
&\quad = \mathbf{B}_0 \cdot \operatorname{curl} \left(\frac{c}{\epsilon} \operatorname{curl} \mathbf{B}_{\nu-1} \right) + B_0 \cdot \left\{ (\mathbf{B}_\nu \cdot \operatorname{grad} \psi) \operatorname{grad} \frac{1}{\epsilon} \right. \\
&\quad \left. - \left(\mathbf{B}_\nu \cdot \operatorname{grad} \frac{1}{\epsilon} \right) \operatorname{grad} \psi \right\}.
\end{aligned}
$$

We now use the vector identity

$$
\operatorname{grad} \left\{ \frac{1}{\epsilon} (\operatorname{grad} \psi \cdot \mathbf{B}_\nu) \right\} = \operatorname{grad} \frac{1}{\epsilon} (\operatorname{grad} \psi \cdot \mathbf{B}_\nu) + \frac{1}{\epsilon} \operatorname{grad} (\operatorname{grad} \psi \cdot \mathbf{B}_\nu)
$$

to replace the last term on the left side of (19′) by its equivalent, cancel terms,

use the fact that $\mathbf{B}_0 \cdot \operatorname{grad} \psi = 0$ to throw out the third term on the left side and the last on the right, and obtain

$$
\mathbf{B}_0 \cdot \left\{ -\frac{2}{\epsilon} \frac{\partial \mathbf{B}_\nu}{\partial \tau} - \mathbf{B}_\nu \operatorname{div} \left(\frac{\operatorname{grad} \psi}{\epsilon} \right) + \frac{1}{\epsilon} \operatorname{grad} (\operatorname{grad} \psi \cdot \mathbf{B}_\nu) \right\}
$$
(20′)
$$
= \mathbf{B}_0 \cdot \operatorname{curl} \left(\frac{c}{\epsilon} \operatorname{curl} \mathbf{B}_{\nu-1} \right).
$$

By (12) we have

$$
\mathbf{B}_0 \cdot \left\{ -\frac{2}{\epsilon} \frac{\partial \mathbf{B}_\nu}{\partial \tau} - \mathbf{B}_\nu \operatorname{div} \left(\frac{\operatorname{grad} \psi}{\epsilon} \right) \right\}
$$
(21′)
$$
= \mathbf{B}_0 \cdot \left\{ \operatorname{curl} \left(\frac{c}{\epsilon} \operatorname{curl} \mathbf{B}_{\nu-1} \right) - \frac{1}{\epsilon} \operatorname{grad} \left(\frac{c}{\mu} \operatorname{div} \mu \mathbf{B}_{\nu-1} \right) \right\}.
$$

Again using $\operatorname{div} (\operatorname{grad} \psi/\epsilon) = (1/\epsilon)\, \Delta_\epsilon \psi$ and multiplying by ϵ we obtain

(22′) $$\mathbf{B}_0 \cdot \left(-2 \frac{\partial \mathbf{B}_\nu}{\partial \tau} - \mathbf{B}_\nu \Delta_\epsilon \psi \right) = \mathbf{B}_0 \cdot \left\{ \epsilon \operatorname{curl} \left(\frac{c}{\epsilon} \operatorname{curl} \mathbf{B}_{\nu-1} \right) - \operatorname{grad} \left(\frac{c}{\mu} \operatorname{div} \mu \mathbf{B}_{\nu-1} \right) \right\}.$$

By carrying the analogous argument thus far we see what the equation for $\mathbf{B}_\nu$ will look like. It is clear from comparing (22) and (22′) that the ordinary differential equation for $\mathbf{B}_\nu$ reads

(49) $$2 \frac{d\mathbf{B}_\nu}{d\tau} + \mathbf{B}_\nu \Delta_\epsilon \psi + 2 \left(\frac{\operatorname{grad} n}{n} \cdot \mathbf{B}_\nu \right) \cdot \operatorname{grad} \psi = -\mathbf{D}_\nu ,$$

where $\mathbf{D}_\nu = \epsilon \operatorname{curl} (c/\epsilon) \operatorname{curl} \mathbf{B}_{\nu-1} - \operatorname{grad} ((c/\mu) \operatorname{div} \mu \mathbf{B}_{\nu-1})$.
Of course the argument required to establish (49) from (22′) must be carried through but the only question, after noting that ϵ replaces μ and μ replaces ϵ in (22′) as against (22), is a possible difficulty with signs since equations (1) and (2), (3) and (4), and (23) and (24) of the text are asymmetric with respect to signs. However these differences offset each other and (49) is the correct equation for $\mathbf{B}_\nu$. We understand that for $\nu = 0$, $\mathbf{C}_\nu = \mathbf{D}_\nu = 0$, for equations (3) and (4) in this case have zero on the right side.

Field Representations in Spherically Stratified Regions*

By NATHAN MARCUVITZ
Polytechnic Institute of Brooklyn

1. *Introduction*

The systematic treatment of electromagnetic radiation and diffraction problems in spherically stratified regions requires the ability to obtain a representation of the vector electromagnetic field produced by prescribed and induced sources in the given region. Such representations may be obtained by a systematization of the classical method of separation of variables termed variously the method of characteristic (or eigen) functions or the method of guided waves (or modes). The desired field representation is expressed as a superposition of mode functions that are so chosen as to permit a simple evaluation of the associated amplitude functions. The problem of finding such a set of modes leads to one or more eigenvalue problems of the Sturm-Liouville type in which the eigenfunctions are characteristic of one or more of the spherical (r, θ, φ) components of the wave operator $-\nabla^2$. These mode functions form a complete set of orthogonal functions in either the r, θ or φ directions and possess in general both a discrete and continuous spectrum. With the determination of such a set of mode functions the original three dimensional field problem may be reduced to a one dimensional problem for the mode amplitudes. The latter is an ordinary differential equation problem characteristic of wave propagation along a single direction—the transmission direction, and is phrased advantageously as a generalized transmission line problem.

The indicated reduction constitutes a "diagonalization" procedure that can be effected in various ways depending on whether one employs a representation in terms of eigenfunction of θ and φ, i.e. waves guided along r, or, if there is no φ dependence, in terms of eigenfunctions in r, waves guided along θ, etc. Each of the resulting field representations has a rate of convergence dependent

---

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

---

*This work was performed at Washington Square College of Arts and Science, New York University and was supported in part by Contract No. AF-19(122)-42 with the U.S. Air Force through sponsorship of the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

upon the parameters involved. In the problem of radiation from an antenna above a spherical earth with various atmospheric conditions, the representation in terms of guided waves along θ is usually the most convergent. All representations are interrelated, however. Starting from a representation with poor convergence, one can obtain by a summation technique the representation with better convergence. It is this latter procedure that has been employed by Watson, van der Pol, Bremmer [1], et al. in their discussion of spherical earth problems. The direct application for such problems of the rapidly convergent representation in waves guided along θ has been made by Booker and Walkinshaw.[1] Although such applications are considered below, our ultimate interest lies in the general representation theory necessary for the solution, via the theory of guided waves, of diffraction problems involving discontinuities in spherical regions.

The general electromagnetic diffraction problem involves the solution of the vector field equations with arbitrary electric and magnetic current sources. The sources are either prescribed or induced (and hence initially unknown), the latter arising from the presence of discontinuities. In the steady state for which a time dependence exp $\{+j\omega t\}$ (or exp $\{-i\omega t\}$) is suppressed, the rms electric field $\mathbf{E}$ and the rms magnetic field $\mathbf{H}$ are determined by

$$\begin{aligned} \nabla \times \mathbf{E} &= -jk\mu\mathbf{H} - \mathbf{M} \\ \nabla \times \mathbf{H} &= jk\epsilon\,\mathbf{E} + \mathbf{J} \end{aligned} \tag{1.1}$$

where $\mathbf{M}(\mathbf{r})$ and $\mathbf{J}(\mathbf{r})$ are respectively the magnetic and electric current densities, $\mu(r)$ and $\epsilon(r)$ (functions of r only) are the relative dielectric constant and the relative permeability, and k is the wave number in vacuum. For mathematical simplicity the normalization of the field quantities has been so chosen that the intrinsic impedance $(\mu_0/\epsilon_0)^{\frac{1}{2}}$ of vacuum is unity. It is desired to obtain in terms of $\mathbf{J}$ and $\mathbf{M}$ a solution to equation (1.1) satisfying such boundary conditions as

$$\mathbf{n} \times \mathbf{E} = 0 = \mathbf{H}\cdot\mathbf{n} \tag{1.2}$$

on perfect metals, where $\mathbf{n}$ is the normal to the metal surface; or such conditions as finiteness and single-valueness in closed but unbounded regions, etc.

In Section 2 the reformulation of the vector problem posed by equations (1.1-2) as ordinary scalar problems of Sturm-Liouville type is considered in detail. There is presented herein the basic mode representation theory necessary for the solution of special and general field problems in spherically stratified regions in terms of guided waves along either the r or θ directions. The determination of the field of an arbitrary current source is reduced hereby to the solution, as a function of the mode index, of an ordinary second order inhomogeneous differential equation, or equivalently, of two simultaneous first order

[1] "Mode Theory of Tropospheric Refraction . . . ." Joint Conference of Phys. Soc. and Royal Meter. Soc., April, 1946, Appendix II. Also H. Bremmer, loc. cit. p. 202–7 and A. Sommerfeld, *Partial Differential Equations*, Academic Press, New York, 1948, p. 214.

(transmission line) equations. In Section 3 a general δ function procedure for the explicit solution of the eigenvalue problems formulated in Section 2 is discussed via the methods of Weyl, Titchmarsh, et al. The interrelation of eigenvalue and transmission line problems in terms of the Green's functions that characterize both types of problem is pointed out. In Section 4 complete sets of mode functions of use in typical spherical problems are evaluated. From a mathematical point of view these complete sets of orthogonal functions provide the basis for symmetric and biorthogonal transform theorems. Several applications of the previous results to the representation of the field of a vertical electric dipole in stratified regions are presented in Section 5.

2a. *Special Field Representations*

In this section we shall be concerned with the representation problem posed by the vector field equations (1.1) and its reduction to a scalar problem of the general Sturm-Liouville type. Let us first consider the special case wherein the excitation is characterized by φ independent, radially directed, electric and magnetic current densities, J_r and M_r. Under these circumstances the vector field equations, when expressed in polar coordinates r, θ, φ, are independent of φ. Hence, they can be separated into the two independent groups:

E type

$$\frac{1}{r \sin\theta}\frac{\partial}{\partial\theta}\sin\theta H_\varphi = jk\epsilon E_r + J_r$$

$$-\frac{1}{r}\frac{\partial}{\partial r} rH_\varphi = jk\epsilon E_\theta \tag{2.1a}$$

$$\frac{1}{r}\frac{\partial}{\partial r} rE_\theta - \frac{1}{r}\frac{\partial}{\partial\theta}E_r = -jk\mu H_\varphi$$

H type

$$\frac{1}{r \sin\theta}\frac{\partial}{\partial\theta}\sin\theta E_\varphi = -jk\mu H_r - M_r$$

$$\frac{1}{r}\frac{\partial}{\partial r} rE_\varphi = jk\mu H_\theta \tag{2.1b}$$

$$\frac{1}{r}\frac{\partial}{\partial r} rH_\theta - \frac{1}{r}\frac{\partial}{\partial\theta}H_r = jk\epsilon E_\varphi .$$

Equations (2.1) can be put into a form wherein transmission along either the r or the θ directions is emphasized.

For the case of r-transmission the transverse (to r) parts of equations (2.1a and b) can be rewritten on elimination of E_r and H_r respectively, as

E type

$$\text{(2.2a)}\qquad \begin{aligned} -\frac{1}{r}\frac{\partial}{\partial r}rE_\theta &= jk\left(\mu + \frac{1}{k^2r^2\epsilon}\frac{\partial}{\partial\theta}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\right)H_\varphi + \frac{1}{jk\epsilon}\frac{\partial J_r}{\partial\theta} \\ -\frac{1}{r}\frac{\partial}{\partial r}rH_\varphi &= jk\epsilon E_\theta \end{aligned}$$

H type

$$\text{(2.2b)}\qquad \begin{aligned} \frac{1}{r}\frac{\partial}{\partial r}rE_\varphi &= jk\mu H_\theta \\ \frac{1}{r}\frac{\partial}{\partial r}rH_\theta &= jk\left[\epsilon + \frac{1}{k^2r^2\mu}\frac{\partial}{\partial\theta}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\right]E_\varphi - \frac{1}{jk\mu}\frac{\partial}{\partial\theta}M_r\,. \end{aligned}$$

The longitudinal components E_r and H_r then follow from the transverse components by the first of equations (2.1a and b), respectively. Differentiability of J_r and M_r with respect to θ is assumed in equations (2.2).

Alternatively, for the case of θ-transmission it is convenient to express the transverse (to θ) parts of equations (2.1) in the form

E type

$$\text{(2.3a)}\qquad \begin{aligned} \frac{1}{r}\frac{\partial}{\partial\theta}E_r &= jk\left[\mu + \frac{1}{k^2r}\frac{\partial}{\partial r}\frac{1}{\epsilon}\frac{\partial}{\partial r}r\right]H_\varphi \\ \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\sin\theta H_\varphi &= jk\epsilon E_r + J_r \end{aligned}$$

H type

$$\text{(2.3b)}\qquad \begin{aligned} -\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\sin\theta E_\varphi &= jk\mu H_r + M_r \\ -\frac{1}{r}\frac{\partial}{\partial\theta}H_r &= jk\left[\epsilon + \frac{1}{k^2r}\frac{\partial}{\partial r}\frac{1}{\mu}\frac{\partial}{\partial r}r\right]E_\varphi \end{aligned}$$

with the longitudinal components E_θ and H_θ determined from the second of equations (2.1a and b), respectively.

$2a_1$. *r-Transmission Formulation*

Explicit solutions of equations (2.2) in a spherical region $r_2 < r < r_1$, $\theta_2 < \theta < \theta_1$ can be obtained in the form of a representation involving ortho-

normal functions of either θ or r. In the former case, now to be discussed, one employs the transverse field representations

E type

$$rE_\theta(r, \theta) = \sum_i V_i'(r)\hat{e}_i'(\theta)$$

(2.4a)

$$rH_\varphi(r, \theta) = \sum_i I_i'(r)\hat{h}_i'(\theta)$$

H type

$$rE_\varphi(r, \theta) = \sum_i V_i''(r)\hat{e}_i''(\theta)$$

(2.4b)

$$rH_\theta(r, \theta) = \sum_i I_i''(r)\hat{h}_i''(\theta)$$

where here and in the following the summation sign signifies either summation over a discrete index i or integration over a continuous index i, or both.

The orthonormal functions $\hat{e}_i'$, $\hat{h}_i'$ are defined in such a manner as to simplify the determination of amplitudes V_i', I_i' from the E type transverse equations (2.2a). The desired simplification is obtained by defining $\hat{h}_i' = \hat{e}_i'$, with the $\hat{h}_i'$ functions determined by the scalar eigenvalue problem

$$\left(\frac{d}{d\theta}\frac{1}{\sin\theta}\frac{d}{d\theta}\sin\theta + \hat{\kappa}'^2\right)\hat{h}_i' = 0, \tag{2.5a}$$

the eigenvalues $\hat{\kappa}_i'^2$ being determined by subjecting the $\hat{h}_i'$ to the boundary conditions

$$\left[H_\varphi \frac{\partial}{\partial\theta}(\sin\theta\hat{h}_i') - h_i'\frac{\partial}{\partial\theta}(\sin\theta H_\varphi)\right]_{\theta_1}^{\theta_2} = 0,^2 \tag{2.5b}$$

which imply that both H_φ and $\hat{h}_i'$ satisfy the same boundary conditions.

Similarly, the orthonormal functions $\hat{e}_i''$, $\hat{h}_i''$ are so defined as to simplify the evaluation of the amplitudes V_i'', I_i'' from the transverse H type equations (2.2b). In this case the desired simplification occurs by defining $\hat{e}_i'' = \hat{h}_i''$, the $\hat{e}_i''$ being determined by the eigenvalue problem

$$\left(\frac{d}{d\theta}\frac{1}{\sin\theta}\frac{d}{d\theta}\sin\theta + \hat{\kappa}''^2\right)\hat{e}_i'' = 0, \tag{2.6a}$$

with the boundary condition

$$\left[E_\varphi \frac{\partial}{\partial\theta}(\sin\theta\hat{e}_i'') - \hat{e}_i''\frac{\partial}{\partial\theta}(\sin\theta E_\varphi)\right]_{\theta_1}^{\theta_2} = 0. \tag{2.6b}$$

The latter imply that both E_φ and $\hat{e}_i''$ satisfy the same boundary conditions.

---

[2] $[\ ]_{\theta_1}^{\theta_2}$ represents the difference between the values of the bracketed quantity at θ_2 and at θ_1.

Equations (2.5) and (2.6) constitute Stürm-Liouville problems whose explicit solution requires the specification of the boundary conditions on the fields E_φ and H_φ ; these conditions must be such as not to destroy the separability into the E and H type equations. For example, if the spherical region under consideration is bounded by two perfectly conducting metallic cones of apertures θ_2 and $\theta_1 (\neq 0, \pi)$

$$E_\varphi = 0 = \frac{\partial}{\partial \theta} (\sin \theta H_\varphi) \qquad \text{at } \theta = \theta_1 \, , \, \theta_2 \tag{2.7a}$$

provided J_r vanishes on the conical boundaries. For "proper" eigenfunctions (cf. Sec. 3) the boundary conditions (2.5b) and (2.6b) then reduce, respectively, to

$$\left.\begin{aligned} \frac{d}{d\theta} (\sin \theta \hat{h}_i') &= 0 \\ \hat{e}_i'' &= 0 \end{aligned}\right\} \qquad \text{at } \theta = \theta_1 \, , \, \theta_2 \, . \tag{2.7b}$$

By a conventional argument it then follows from equations (2.5-6) that both the eigenfunctions $\hat{e}_i'$ and $\hat{e}_i''$ are orthogonal and can be so normalized as to possess the properties (cf. Sec. 3)

$$\int_{\theta_1}^{\theta_2} \hat{e}_i(\theta) \hat{e}_j(\theta) \sin \theta \, d\theta = \delta_{ij} \, .^3 \tag{2.8}$$

As a further example consider the case $\theta_1 = 0$, $\theta_2 = \pi$ corresponding to an unbounded spherical region. In this singular case the boundary conditions on the fields are, in view of the finiteness of the latter,

$$\frac{\partial}{\partial \theta} (\sin \theta E_\varphi) = 0 = \frac{\partial}{\partial \theta} (\sin \theta H_\varphi) \qquad \text{at } \theta = 0, \pi. \tag{2.9a}$$

The associated boundary conditions (2.5b) and (2.6b) on the "proper" eigenfunctions are then

$$\left.\begin{aligned} \frac{d}{d\theta} (\sin \theta \hat{h}_i') &= 0 \\ \frac{d}{d\theta} (\sin \theta \hat{e}_i'') &= 0 \end{aligned}\right\} \qquad \text{at } \theta = 0, \pi \tag{2.9b}$$

from which by means of equations (2.5-6) the orthogonality properties (2.8) with $\theta_1 = 0$, $\theta_2 = \pi$ likewise follow.

The format and notation employed in the representations (2.4) and the

---

[3] $$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

eigenvalue problems (2.5-6) although applied to single-component, i.e. scalar, representations are capable of generalization to the two-component vector representations necessary when $\partial/\partial\varphi \neq 0$. Since the scalar nature of the E- and H-type representation can still be retained in the latter case (cf. Sec. 2b), it is desirable to rephrase the eigenvalue problems of (2.5-6) in accord with the format for the general case. One introduces scalar mode functions ϕ_i and ψ_i by the detailed first order equations

$$E \textit{ type } (\hat{e}' = \hat{h}')$$

$$\frac{d}{d\theta}\phi_i = -\hat{\kappa}_i'\hat{e}_i' \tag{2.10a}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}(\sin\theta\hat{e}_i') = \hat{\kappa}_i'\phi_i$$

$$H \textit{ type } (\hat{e}'' = \hat{h}'')$$

$$\frac{d}{d\theta}\psi_i = -\hat{\kappa}_i''\hat{h}_i'' \tag{2.10b}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}(\sin\theta\hat{h}_i'') = \hat{\kappa}_i''\psi_i .$$

On elimination of ϕ_i and ψ_i it is evident that equations (2.10a and b) are completely equivalent to equations (2.5a) and (2.6a). Alternatively, on elimination of $\hat{e}_i'$ and $\hat{h}_i''$ defining equations for ϕ_i and ψ_i can be obtained in the form of the second order equations

$$\left(\frac{1}{\sin\theta}\frac{d}{d\theta}\sin\theta\frac{d}{d\theta} + \hat{\kappa}_i'^2\right)\phi_i = 0 \tag{2.10c}$$

$$\left(\frac{1}{\sin\theta}\frac{d}{d\theta}\sin\theta\frac{d}{d\theta} + \hat{\kappa}_i''^2\right)\psi_i = 0 \tag{2.10d}$$

which admit unique solutions when ϕ_i and ψ_i are subject to suitable boundary conditions. For the case contemplated in (2.7b) these conditions reduce for the "proper" eigenfunctions to

$$\phi_i = 0,^4 \qquad \frac{d\psi_i}{d\theta} = 0 \qquad \text{at } \theta = \theta_{1,2} ,$$

whereas the conditions (2.9b) require that ϕ_i and ψ_i be finite at the singular points $\theta = 0, \pi$. From (2.8) or directly from (2.10c, d) it then follows that the ϕ_i and ψ_i possess the orthonormality properties

[4]This condition must be modified when $\hat{\kappa}_i' = 0$. This latter possibility arises in coaxial structures and characterizes the so-called principal mode. For this mode the condition $\phi_i = 0$ is replaced by a condition of constancy of $\phi_i/\hat{\kappa}_i$ on the various peripheries of the cross-section.

$$\int_{\theta_1}^{\theta_2} \phi_i(\theta)\phi_j(\theta) \sin \theta \, d\theta = \delta_{ij} \,. \tag{2.11}$$

The solution of the scalar eigenvalue problems (2.10c and d) together with the relations (2.10a and b) usually provide the simplest procedure for determining the mode functions $\hat{e}_i'$ and $\hat{e}_i''$.

In view of the orthogonality properties (2.8) of the mode functions the V_i, I_i amplitudes in (2.4) can be readily expressed in terms of the fields by relations of the form (omitting the distinguishing superscripts)

$$\begin{aligned} V_i(r) &= \int_{\theta_1}^{\theta_2} rE(r, \theta)e_i(\theta) \sin \theta \, d\theta \\ I_i(r) &= \int_{\theta_1}^{\theta_2} rH(r, \theta)h_i(\theta) \sin \theta \, d\theta. \end{aligned} \tag{2.12}$$

The defining equations for $V_i(r)$ and $I_i(r)$ can be obtained by transformation of equations (2.2) in accordance with the operations indicated in (2.12). Thus, multiplying the E-type equations (2.2a) by $\hat{e}_i'$ sin θ and the H-type equations (2.2b) by $\hat{e}_i''$ sin θ and integrating over θ from θ_1 to θ_2, one obtains on use of equations (2.5,6, and 12)

$$\begin{aligned} -\frac{dV_i}{dr} &= j\kappa_i Z_i I_i + v_i \\ -\frac{dI_i}{dr} &= j\kappa_i Y_i V_i + i_i \end{aligned} \tag{2.13a}$$

where

$$\kappa_i = (k^2\epsilon\mu - \hat{\kappa}_i^2/r^2)^{1/2};$$

for the E-modes

$$Z_i' = \frac{1}{Y_i'} = \frac{\kappa_i'}{k\epsilon}$$

$$v_i' = v_i'(r) = -Z_i' \int_{\theta_1}^{\theta_2} rJ_r(r, \theta)e_{ri}(\theta) \sin \theta \, d\theta \tag{2.13b}$$

$$i_i' = 0$$

whereas for the H-modes

$$Z_i'' = (1/Y_i'') = (k\mu/\kappa_i'')$$

$$v_i'' = 0 \tag{2.13c}$$

$$i_i'' = i_i''(r) = -Y_i'' \int_{\theta_1}^{\theta_2} rM_r(r, \theta)h_{ri}(\theta) \sin \theta \, d\theta$$

The superscripts distinguishing the mode type have been omitted in equations (2.13a) since the equations have the same form for all modes. The functions $\hat{e}_{ri}(\theta)$ and $\hat{h}_{ri}(\theta)$ are defined by the relations

$$
\text{(2.13d)} \qquad
\begin{aligned}
j\kappa_i'\hat{e}_{ri} &= \kappa_i'\phi_i = \frac{1}{r\sin\theta}\frac{d}{d\theta}\sin\theta\hat{e}_i' \\
j\kappa_i''\hat{h}_{ri} &= \kappa_i''\psi_i = \frac{1}{r\sin\theta}\frac{d}{d\theta}\sin\theta\hat{h}_i'' \ ;
\end{aligned}
$$

their significance is evident when it is noted by equations (2.1) and (2.4) that the r components of the field may be represented as

$$
E_r = \sum_i Z_i'I_i'(r)e_{ri}(\theta)
$$

$$
H_r = \sum_i Y_i''V_i''(r)h_{ri}(\theta)
$$

in those regions wherein $J_r = 0 = M_r$.

Equations (2.13a) being of conventional transmission line form, constitute the basis for terming the amplitudes V_i and I_i mode voltages and currents, respectively; the inhomogeneous terms v_i and i_i , characteristic of the excitation of the i-th mode, are correspondingly designated as the source voltage and source current, respectively. The indicated variability with r of the propagation wave number κ_i and of the characteristic impedance Z_i implies that equations (2.13a) are spherical transmission line equations characteristic of the r variation of spherical waves. The wave character is made more explicit by casting (2.13a) in the form of second order equations. Thus, eliminating V_i' from (2.13a), one obtains for the case of the E-mode wave equation

$$
\text{(2.14a)} \qquad \left(\frac{d}{dr}\frac{1}{\epsilon}\frac{d}{dr} + k^2\mu - \frac{\hat{\kappa}_i'^2}{\epsilon r^2}\right)I_i' = -jkv_i'
$$

whereas eliminating I_i'' from (2.13a), one finds for the H-mode wave equation

$$
\text{(2.14b)} \qquad \left(\frac{d}{dr}\frac{1}{\mu}\frac{d}{dr} + k^2\epsilon - \frac{\hat{\kappa}_i''^2}{\mu r^2}\right)V_i'' = -jki_i'' \ .
$$

The amplitudes V_i' and I_i'' readily follow by (2.13a) from the solutions I_i' and V_i'' of (2.14). Explicit solutions of the inhomogeneous equations (2.14) will be discussed in Sections 3-4.

With the knowledge of the eigenfunctions $\hat{e}_i$ and eigenvalues $\hat{\kappa}_i^2$ from (2.5-6) the problem of finding the solution of the partial differential equations (2.2) in the form of the representation (2.4) is reduced to that of solving the set of ordinary modal equations given in (2.13) or (2.14). Since the solution for one mode is typical for every other, the latter problem is solved as a function of the mode index i. Although the representation so obtained constitutes a formal

solution, its practical usefulness is dependent on the rapidity of convergence of (2.4).

2a$_2$. *θ-Transmission Formulation*

Explicit solutions of the φ independent field equations (2.2) in a spherical region $r_2 < r < r_1$, $\theta_2 < \theta < \theta_1$ can equally well be obtained by employing a representation involving orthonormal functions of r. However, the resulting θ-transmission formalism does not possess the same general character as that for r-transmission because of the lack of vector separability of the field equations in directions transverse to θ. For the θ-transmission development one starts with the (transverse to θ) field representations

E type

$$r^2E_r(r,\ \theta) = \sum_i V_i'(\theta)e_i'(r)$$

(2.15a)

$$rH_\varphi(r,\ \theta) = \sum_i \frac{I_i'(\theta)}{\sin\theta}\, h_i'(r)$$

H type

$$r^2H_r(r,\ \theta) = \sum_i \hat{I}_i''(\theta)h_i''(r)$$

(2.15b)

$$rE_\varphi(r,\ \theta) = \sum_i \frac{V_i''(\theta)}{\sin\theta}\, e_i''(r).$$

The orthonormal functions e_i' , h_i' in (2.15a) are so defined as to simplify the evaluation of the amplitudes $\hat{V}_i'$, $\hat{I}_i'$ from the transverse equations (2.3a). The desired simplification is achieved by defining $h_i' = -\epsilon e_i'$ with the functions h_i' determined by

$$\left(\frac{d}{dr}\frac{1}{\epsilon}\frac{d}{dr} + k^2\mu - \frac{\hat{\kappa}_i^2}{\epsilon r^2}\right)h_i' = 0 \tag{2.16a}$$

and the boundary conditions[5]

$$\left[\frac{1}{\epsilon}\left(h_i'\frac{\partial}{\partial r}rH_\varphi - rH_\varphi\frac{\partial}{\partial r}h_i'\right)\right]_{r_1}^{r_2} = 0. \tag{2.16b}$$

Equations (2.16) constitute an eigenvalue problem whose solution yields a set of orthogonal eigenfunctions h_i' and associated eigenvalues $\hat{\kappa}_i'^2$. Conditions (2.16b) apply when ϵ is a continuous function of r. For discontinuous ϵ supplemental conditions of continuity of $(1/\epsilon)$ (d/dr) h_i' and h_i' are also necessary, the latter being a consequence of the continuity of $(1/\epsilon)$ $(\partial/\partial r)rH_\varphi$ in spherically stratified regions. Equations (2.16b) imply that the h_i' obey the same boundary conditions as H_φ .

[5] $[\ ]_{r_1}^{r_2}$ represents the difference between the values of the bracketed quantity at r_2 and at r_1 .

The orthonormal functions e_i'', h_i'' in (2.15b) are defined so as to permit a simple evaluation of the amplitudes $\hat{V}_i''$, $\hat{I}_i''$ from equations (2.3b). In this case simplification results if one defines $\mu h_i'' = e_i''$, with the mode functions e_i'' defined by the equations

$$\left(\frac{d}{dr}\frac{1}{\mu}\frac{d}{dr} + k^2\epsilon - \frac{\hat{\kappa}_i''^2}{\mu r^2}\right)e_i'' = 0 \tag{2.17a}$$

and the boundary conditions

$$\left[\frac{1}{\mu}\left(e_i''\frac{\partial}{\partial r}rH_\varphi - rE_\varphi\frac{\partial}{\partial r}e_i''\right)\right]_{r_1}^{r_2} = 0. \tag{2.17b}$$

The solution of this eigenvalue problem yields a set of orthogonal eigenfunctions e_i'' and eigenvalues $\hat{\kappa}_i''^2$. In analogy with (2.16b) the boundary conditions (2.17b) on the e_i'' are applicable only when μ is a continuous function of r. For discontinuous μ there are additional conditions of continuity on $(1/\mu)\ (d/dr)\ e_i''$ and e_i'' at the discontinuity points.

The general Sturm-Liouville problems defined in (2.16) and (2.17) lead to unique sets of orthonormal eigenfunctions only on specification of the boundary conditions on the fields E_φ and H_φ. For example, if the spherical region under consideration is characterized by continuous ϵ, μ and is bounded by perfectly conducting spherical segments at r_1 and r_2,

$$E_\varphi = 0 = (\partial/\partial r)rH_\varphi \qquad \text{at } r = r_1\,,\ r_2 \tag{2.18a}$$

Hence, conditions (2.16b) and (2.17b) reduce for "proper" eigenfunctions (cf. Sec. 3) to, respectively,

$$(d/dr)h_i' = 0 \qquad \text{and} \qquad e_i'' = 0 \qquad \text{at } r = r_1 r_2\,. \tag{2.18b}$$

Alternatively, the spherical region may be open at both ends—i.e. $r_1 = 0$ and $r_2 = \infty$. The latter are singular points of the differential equations (2.16a) and (2.17a) and hence the dependence of the solutions on the boundary conditions is somewhat different than at regular points. Weyl [2] has distinguished between two kinds of singular points: one yielding solutions of the "limit point" type, the other of "limit circle" type. Loosely stated, these terms imply an independence of the solutions on the boundary conditions in the former case, and a regular dependence in the latter. Since $r = 0$ corresponds to the limit point case, the boundary condition at $r = 0$ may be stated simply as a condition of finiteness of the mode functions and their r derivatives. On the other hand $r = \infty$ corresponds to the limit circle case and hence the boundary conditions at $r = \infty$ must be more definite. Since the fields of physical interest usually satisfy the "radiation conditions"

$$\left[\frac{\partial}{\partial r} + jk(\mu\epsilon)^{1/2}\right]_{rH_\varphi}^{rE_\varphi} \to 0 \qquad \text{as } r \to \infty \tag{2.19a}$$

the corresponding conditions on the "proper" mode functions are by (2.16b) and (2.7b)

$$\left[\frac{d}{dr} + jk(\mu\epsilon)^{1/2}\right]_{e_i''}^{h_i'} \to 0 \qquad \text{as } r \to \infty \tag{2.19b}$$

assuming that μ, ϵ are constant in the limit $r \to \infty$.

With the specification of the boundary conditions on the fields the Sturm-Liouville problems in (2.16) and (2.17) become completely defined. Although the explicit solution of such problems will be deferred until Section 3, it is desirable to point out at this point the orthogonality properties of the mode functions. Thus for the case contemplated by the boundary conditions (2.18), it follows from (2.16) and (2.17) by a conventional argument that

$$-\int_{r_1}^{r_2} e_i'(r)h_j'(r)\,\frac{dr}{r^2} = \delta_{ij} \tag{2.20a}$$

and

$$\int_{r_1}^{r_2} e_i''(r)h_j''(r)\,\frac{dr}{r^2} = \delta_{ij} \tag{2.20b}$$

plus the other forms obtained from the identities $h' = -\epsilon e'$ and $\mu h'' = e''$. In view of the orthonormality properties (2.20) the $\hat{V}_i$, $\hat{I}_i$ amplitudes in the representation (2.15) can be expressed in terms of the fields as

$$\hat{V}_i'(\theta) = -\int_{r_1}^{r_2} E_r(r, \theta)h_i'(r)\,dr; \qquad \frac{I_i'(\theta)}{\sin\theta} = -\int_{r_1}^{r_2} H_\varphi(r, \theta)e_i'(r)\,\frac{dr}{r} \tag{2.21a}$$

and

$$I_i''(\theta) = \int_{r_1}^{r_2} H_r(r, \theta)e_i''(r)\,dr; \qquad \frac{V_i''(\theta)}{\sin\theta} = \int_{r_1}^{r_2} E_\varphi(r, \theta)h_i''(r)\,\frac{dr}{r}\,. \tag{2.21b}$$

The defining equations for the mode amplitudes V_i, I_i can now be determined from the transverse fields equations (2.3) by utilizing the transformation relations (2.21). Thus, multiplying the E-type equations (2.3a) by $\epsilon\, e_i'r$ or e_i' and the H type equations (2.3b) by h_i'' or $\mu\, h_i''\, r$ and integrating over r from r_1 to r_2, one obtains

$$\begin{aligned} -\frac{d\hat{V}_i}{d\theta} &= j\hat{\kappa}_i\hat{Z}_i\hat{I}_i + \hat{v}_i(\theta) \\ -\frac{d\hat{I}_i}{d\theta} &= j\hat{\kappa}_i\hat{Y}_i\hat{V}_i + \hat{\imath}_i(\theta) \end{aligned} \tag{2.22a}$$

where for E-type modes

$$\hat{Z}_i' = \frac{1}{\hat{Y}_i'} = \frac{\hat{\kappa}_i'}{k \sin \theta}$$

(2.22b)
$$\hat{v}_i'(\theta) = 0$$

$$\hat{\imath}_i'(\theta) = \sin \int_{r_1}^{r_2} J_r(r, \theta) e_i'(r)\, dr$$

and for H-type modes

$$\hat{Z}_i'' = \frac{1}{\hat{Y}_i''} = \frac{k \sin \theta}{\hat{\kappa}_i''}$$

(2.22c)
$$\hat{v}_i''(\theta) = \sin \theta \int_{r_1}^{r_2} M_r(r, \theta) h_i''(r)\, dr$$

$$\hat{\imath}_i''(\theta) = 0.$$

Equations (2.22a) have the form of transmission line equations, one for each mode. The indicated variability with θ of the characteristic impedances $\hat{Z}_i$ implies that (2.22a) describe wave propagation on "angular" transmission lines whose propagation wave numbers $\hat{\kappa}_i$ are determined by the eigenvalue problems in (2.16-7). As in equations (2.13) $\hat{V}_i$, $\hat{I}_i$ are designated as mode voltages and currents, while $\hat{v}_i$ and $\hat{\imath}_i$ are termed source voltages and currents. It is frequently convenient to cast the first order transmission line equations in the form of second order wave equations. Thus, by elimination of $\hat{I}_i'$ from (2.22a), one has for the E-type modes

(2.23a)
$$\left(\frac{1}{\sin \theta} \frac{d}{d\theta} \sin \theta \frac{d}{d\theta} + \hat{\kappa}_i'^2\right) \hat{V}_i' = -j\hat{\kappa}_i' \hat{Z}_i' \hat{\imath}_r(\theta),$$

and for the H-type modes by eliminating $\hat{V}_i''$

(2.23b)
$$\left(\frac{1}{\sin \theta} \frac{d}{d\theta} \sin \theta \frac{d}{d\theta} + \hat{\kappa}_i''^2\right) \hat{I}_i'' = -j\hat{\kappa}_i'' \hat{Y}_i'' \hat{v}_i(\theta).$$

The knowledge of the mode functions e_i and of the eigenvalues $\hat{\kappa}_i^2$ reduces the problem of finding solutions of the partial differentials equations (2.3) to that of solving the ordinary differential equation problems posed in (2.22) or (2.23). The analysis for one mode is typical of that for all the others; the desired solution in the form (2.15) is then found by synthesis of the modal solutions.

$2a_3$. *Potential Formulations*

A procedure, alternative but intimately related to those discussed above, for obtaining a solution of the φ independent field equations (2.1) employs a representation not of the fields themselves but rather of potential functions from which the fields are derivable. In this so-called method of Hertz (or

Debye) potentials [4] the scalar nature of the representation is introduced at an early stage. The method will be treated quite briefly at this point since it is taken up in more detail in Section 2*b*. For the *E*-type case a potential function $\Pi'(r, \theta)$ is introduced by expressing H_φ by

$$rH_\varphi = -\frac{\partial}{\partial\theta}\Pi'. \tag{2.24}$$

By equations (2.1a) one then finds

$$\begin{aligned} rE_\theta &= \frac{1}{jk\epsilon}\frac{\partial^2}{\partial\theta\,\partial r}\Pi' \\ E_r &= \frac{1}{jk}\left(k^2\mu + \frac{\partial}{\partial r}\frac{1}{\epsilon}\frac{\partial}{\partial r}\right)\Pi' \end{aligned} \tag{2.25a}$$

with Π' given by

$$\left(\frac{\partial}{\partial r}\frac{1}{\epsilon}\frac{\partial}{\partial r} + k^2\mu + \frac{1}{\epsilon r^2}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}\right)\Pi' = -\frac{J_r}{\epsilon} \tag{2.25b}$$

and subject to boundary conditions that follow by (2.24) and (2.25a) from corresponding conditions on the fields.

Representations of $\Pi'(r, \theta)$ can be obtained in terms of orthonormal functions of either r or θ. In the former case one employs the representation

$$\Pi'(r, \theta) = \sum_i I_i'(r)\frac{\phi_i(\theta)}{\hat{\kappa}_i'} \tag{2.26}$$

where the $\phi_i(\theta)$ and $\hat{\kappa}_i'$ are the eigenfunctions and eigenvalues defined in equations (2.10c). In view of the orthonormality properties (2.11) of the ϕ_i one has

$$I_i'(r) = \hat{\kappa}_i'\int_{\theta_1}^{\theta_2}\Pi'(r, \theta)\phi_i(\theta)\sin\theta\,d\theta. \tag{2.27}$$

To determine explicitly the amplitudes I_i' one multiplies equation (2.25b) by $\hat{\kappa}_i'\phi_i\sin\theta$ and integrates over θ from θ_1 to θ_2. One then obtains by (2.10c), (2.27), and the boundary conditions on ϕ_i expressed in the form

$$\left[\sin\theta\left(\phi_i\frac{\partial\Pi'}{\partial\theta} - \Pi'\frac{\partial}{\partial\theta}\phi_i\right)\right]_{\theta_1}^{\theta_2} = 0, \tag{2.28}$$

the modal equations

$$\left(\frac{d}{dr}\frac{1}{\epsilon}\frac{d}{dr} + k^2\mu - \frac{\hat{\kappa}_i'^2}{\epsilon r^2}\right)I_i' = -\frac{\hat{\kappa}_i'}{\epsilon}\int_{\theta_1}^{\theta_2}J_r\phi_i\sin\theta\,d\theta \tag{2.29}$$

for the amplitudes I_i'. Equation (2.29) is seen to be identical with equation (2.14a) previously determined.

Alternatively $\Pi'(r, \theta)$ can be represented in terms of functions orthonormal in the interval $r_1 < r < r_2$ by

$$\Pi'(r, \theta) = \sum_i \hat{V}_i'(\theta) \frac{h_i'(r)}{\hat{\kappa}_i'^2/jk} \cdot \tag{2.30}$$

The mode functions h_i' are the same as those defined in (2.16a). From the orthonormality properties (2.20a) of the functions $h_i' = -\epsilon\, e_i'$ the mode amplitudes $\hat{V}_i'$ follow as

$$\hat{V}_i'(\theta) = \frac{\hat{\kappa}_i'^2}{jk} \int_{r_1}^{r_2} \Pi'(r, \theta) \frac{h_i'(r)\, dr}{\epsilon r^2} \cdot \tag{2.31}$$

In accordance with (2.31), multiplication of (2.25b) by $h_i'\hat{\kappa}_i'^2/jk$, integration from r_1 to r_2, use of (2.16a) and the boundary conditions on h_i' in the form

$$\left[\frac{1}{\epsilon}\left(h_i' \frac{\partial \Pi'}{\partial r} - \Pi' \frac{\partial h_i'}{\partial r}\right)\right]_{r_1}^{r_2} = 0 \tag{2.32}$$

yields as the defining equations for $\hat{V}_i'$

$$\left(\frac{1}{\sin\theta}\frac{d}{d\theta}\sin\theta\frac{d}{d\theta} + \hat{\kappa}_i'^2\right)\hat{V}_i' = \frac{\hat{\kappa}_i^2}{jk}\int_{r_1}^{r_2} J_r e_i'\, dr. \tag{2.33}$$

Equation (2.33) is identical with the previously obtained equation (2.23a).

Since the H-type equations (2.1b) follow from the E-type equations (2.1a) on the duality replacements $E_r \to H_r$, $E_\theta \to H_\theta$, $H_\varphi \to E_\varphi$, $J_r \to M_r$, $\epsilon \to \mu$, and $\mu \to \epsilon$, it is unnecessary to repeat the details of the above formalism for the H-type case. One has only to introduce into equations (2.24-33) the duality replacements $\Pi' \to \Pi''$, $I_i' \to V_i''$, $V_i' \to I_i''$, $\phi_i \to \psi_i$, $\hat{\kappa}_i' \to \hat{\kappa}_i''$, $h_i' \to -e_i''$, and $e_i' \to h_i''$ to obtain the H-type formalism.

2b. *General Field Representation*

The solution of the field equations (1.1) in an r-stratified spherical region with *arbitrary* excitation $\mathbf{J}$ and $\mathbf{M}$ is facilitated by elimination of either the r or θ field components. Only the elimination of the r-components, corresponding to an r-transmission analysis, will be considered here. In this case the resulting transverse equations may be cast in an invariant (two) vector form that can be obtained by vector and scalar product multiplication of equations (1.1) by the radial vector $\mathbf{r}$. These operations yield, respectively,

$$\begin{aligned} \nabla(\mathbf{r}\cdot\mathbf{E}) - (\mathbf{r}\cdot\nabla)\mathbf{E} - \mathbf{E} &= -jk\mathbf{r}\times\mathbf{H} - \mathbf{r}\times\mathbf{M} \\ \nabla(\mathbf{r}\cdot\mathbf{H}) - (\mathbf{r}\cdot\nabla)\mathbf{H} - \mathbf{H} &= jk\mathbf{r}\times\mathbf{E} + \mathbf{r}\times\mathbf{J} \end{aligned} \tag{2.34a}$$

and

$$\text{(2.34b)}\qquad \begin{aligned} \nabla\cdot\mathbf{r}\times\mathbf{E} &= jk\mu\mathbf{r}\cdot\mathbf{H}+\mathbf{r}\cdot\mathbf{M} \\ \nabla\cdot\mathbf{H}\times\mathbf{r} &= jk\epsilon\mathbf{r}\cdot\mathbf{E}+\mathbf{r}\cdot\mathbf{J} \end{aligned}$$

on use of simple vector relations and the fact that $\nabla\times\mathbf{r}=0$. On elimination of the radial field components from (2.34a) by means of (2.34b), the transverse equations follow, as

$$\text{(2.35)}\qquad \begin{aligned} -\frac{\partial}{\partial r}r\mathbf{E}_t &= jk\left(\mu+\frac{1}{k^2\epsilon}\,{}_t\nabla\nabla_t\right)\cdot\mathbf{H}\times\mathbf{r}+\mathbf{M}\times\mathbf{r}+\frac{{}_t\nabla(\mathbf{r}\cdot\mathbf{J})}{jk\epsilon} \\ -\frac{\partial}{\partial r}r\mathbf{H}_t &= jk\left(\epsilon+\frac{1}{k^2\mu}\,{}_t\nabla\nabla_t\right)\cdot\mathbf{r}\times\mathbf{E}_t+\mathbf{r}\times\mathbf{J}+\frac{{}_t\nabla(\mathbf{r}\cdot\mathbf{M})}{jk\mu} \end{aligned}$$

where the subscript t denotes vector components transverse to the $\mathbf{r}$ direction. In accordance with this notation the vector gradient operator ∇ has been decomposed either as

$$\text{(2.36a)}\qquad \nabla=\nabla_t+\frac{1}{r^2}\frac{\partial}{\partial r}r^2\mathbf{r}_0=\nabla-(\nabla\cdot\mathbf{r}_0)\mathbf{r}_0$$

or as

$$\text{(2.36b)}\qquad \nabla={}_t\nabla+\mathbf{r}_0\frac{\partial}{\partial r}=\nabla-\mathbf{r}_0(\mathbf{r}_0\cdot\nabla).$$

The necessity of distinguishing between the transverse component ∇_t taken from the right and ${}_t\nabla$ the transverse component from the left is a consequence of the variability in direction of the unit radial vector $\mathbf{r}_0$.

To evaluate in scalar terms the transverse fields defined by the vector partial differential equations (2.35) one can proceed in either of two ways. The vector fields themselves can be represented as a superposition of an infinite number of characteristic vector modes, each vector mode being then decomposed into two components E and H modes whose scalar amplitudes are determined by ordinary differential equations. Alternatively, the transverse vector fields can be represented in terms of two scalar potentials that in turn can each be represented as a superposition of an infinite number of characteristic scalar modes whose amplitudes are also determined by the same ordinary differential equations. The former procedure was adopted in subsection 1 of Section 2a, the latter in subsection 3 of Section 2a; for the special φ independent excitation J_r and M_r treated therein it was possible to effect a natural separation into the component E and H-modes at all stages of the development. The latter procedure will be employed throughout most of this section in a rather general form; it is essentially the method of Hertz or Debeye potentials and is intimately related to the Green's function techniques discussed in Section 3, et seq.

Any transverse vector can be decomposed into a transverse gradient and

a transverse curl part. Thus, to obtain the desired scalar reformulation of equations (2.35), let

$$(2.37)\qquad \begin{aligned} \mathbf{E}_t &= -{}_t\nabla V' - {}_t\nabla \times V''\mathbf{r}_0\,, & \mathbf{J}_t &= -{}_t\nabla J' - {}_t\nabla \times J''\mathbf{r}_0 \\ \mathbf{H}_t &= -{}_t\nabla I'' + {}_t\nabla \times I'\mathbf{r}_0\,, & \mathbf{M}_t &= -{}_t\nabla M'' + {}_t\nabla \times M'\mathbf{r}_0 \end{aligned}$$

where the scalar functions (of r, θ, φ) V', V'' are constitutive measures of $\mathbf{E}_t$; I', I'' of $\mathbf{H}_t$; etc. The scalar functions I' and V'' appear as generalizations of the potentials Π' and Π'', respectively, of Section 2a. On substituting (2.37) into (2.35), noting that the operators $r_t\nabla$ and $\partial/\partial r$ commute, and equating the independent transverse gradient and curl terms, one obtains as the defining equations for the potentials, the set of scalar equations

E type

$$(2.38\text{a})\qquad \begin{aligned} -\frac{\partial}{\partial r} V' &= jk\left(\mu + \frac{\nabla_t \cdot {}_t\nabla}{k^2\epsilon}\right)I' + M' - \frac{J_r}{jk\epsilon} \\ -\frac{\partial}{\partial r} I' &= jk\epsilon\, V' + J' \end{aligned}$$

and *H type*

$$(2.38\text{b})\qquad \begin{aligned} -\frac{\partial V''}{\partial r} &= jk\mu I'' + M'' \\ -\frac{\partial I''}{\partial r} &= jk\left(\epsilon + \frac{\nabla_t \cdot {}_t\nabla}{k^2\mu}\right)V'' + J'' - \frac{M_r}{jk\mu} \end{aligned}$$

where $\nabla_t \cdot {}_t\nabla$ is the two dimensional operator that in polar coordinates θ, ϕ is given by (2.36) as

$$(2.39)\qquad r^2\nabla_t \cdot {}_t\nabla = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\,\cdot$$

The solutions of the simultaneous partial differential equations (2.38a) and (2.38b) are to be subjected to as yet unspecified boundary conditions on V', I' and V'', I''. For example for a perfectly conducting metallic boundary limiting the cross-sectional region $\theta_1 < \theta < \theta_2$ and $\varphi_1 < \varphi < \varphi_2$, it follows by (1.2) and (2.37) that the boundary conditions are

$$(2.40)\qquad \begin{aligned} V' &= 0 = I' \\ \frac{\partial V''}{\partial \nu} &= 0 = \frac{I''}{\partial \nu} \end{aligned}$$

where $\boldsymbol{\nu}$ is normal to the boundary and perpendicular to the $\mathbf{r}$ direction. For an unbounded θ, φ cross-section the boundary conditions on V', I' and V'',

I'' are simply finiteness and single-valuedness. The cross-sectional boundary conditions are subject to restrictions that make valid the E and H decomposition implied in equations (2.38a and b); this decomposition is valid for the boundary conditions just mentioned. The remaining longitudinal boundary conditions are determined by "impedance" conditions in the r direction. For example, at a perfectly conducting surface with specified r

$$V' = 0 = V''$$

$$\frac{\partial I'}{\partial r} = 0 = \frac{\partial I''}{\partial r};$$

for a radiation condition as $r \to \infty$

$$\frac{V'}{I'} = \left(\frac{\mu}{\epsilon}\right)^{1/2} = \frac{V''}{I''}$$

which implies that there are no sources and ϵ, μ are constant as $r \to \infty$.

The partial differential equations (2.38a) can be reduced to a set of ordinary differential equations by employing a representation in orthonormal functions characteristic of the operator $r^2\, \nabla_t \cdot {}_t\nabla$. Thus let

$$\begin{aligned} V'(r, \theta, \varphi) &= \sum_i V_i'(r) \frac{\phi_i(\theta, \varphi)}{\hat{\kappa}_i'} \\ I'(r, \theta, \varphi) &= \sum_i I_i'(r) \frac{\phi_i(\theta, \varphi)}{\hat{\kappa}_i'} \end{aligned} \tag{2.41}$$

where, in view of (2.38a) the characteristic mode functions ϕ_i are defined by the eigenvalue problem

$$(r^2 \nabla_t \cdot {}_t\nabla + \hat{\kappa}_i'^2)\phi_i = 0 \tag{2.42}$$

with the eigenvalues $\hat{\kappa}_i'^2$ determined by subjecting ϕ_i to appropriate boundary conditions. The mode functions ϕ_i defined in (2.42) possess by the usual argument the orthogonality properties

$$\iint \phi_i \phi_j \, d\Omega = \delta_{ij} \tag{2.43}$$

where the surface integral with respect to $d\Omega = \sin\theta\, d\theta\, d\varphi$ is extended over the spherical cross section (perpendicular to $\mathbf{r}$) of the given region. From (2.41) and (2.43) it follows that the mode amplitudes are given in terms of $V'(r, \theta, \varphi)$ and $I'(r, \theta, \varphi)$ by

$$\begin{aligned} V_i'(r) &= \hat{\kappa}_i' \iint V'\phi_i \, d\Omega \\ I_i'(r) &= \hat{\kappa}_i' \iint I'\phi_i \, d\Omega. \end{aligned} \tag{2.44}$$

To obtain the defining equations for the mode amplitudes by transformation of equations (2.38a) in accordance with (2.44), it is necessary to employ Green's theorem for $\phi_i(\theta, \varphi)$ and $I'(r, \theta, \varphi)$ in the two dimensional form appropriate for a spherical surface, namely:

$$\iint [\phi_i \nabla_t \cdot {}_t\nabla I' - I' \nabla_t \cdot {}_t\nabla \phi_i] r^2 \, d\Omega = \int \left[\phi_i \frac{\partial I'}{\partial \nu} - I' \frac{\partial \phi_i}{\partial \nu} \right] ds. \tag{2.45}$$[6]

The left hand surface integral is extended over the spherical cross section of the given region, whereas the right hand line integral with respect to ds is taken over the peripheral curve s, if any, bounding the cross-section; as before $\boldsymbol{\nu}$ denotes the outward normal direction at s. Equation (2.45) assumes a particularly simple form if the boundary conditions on the ϕ_i are so chosen that

$$\left[\phi_i \frac{\partial I'}{\partial \nu} - I' \frac{\partial \phi_i}{\partial \nu} \right] = 0 \qquad \text{on } s. \tag{2.46}$$

Multiplication of equations (2.38a) by ϕ_i and integration over the cross-section then yields by (2-44-6) as the defining equations for the E-mode amplitudes

$$\begin{aligned} -\frac{dV_i'}{dr} &= j\hat{\kappa}_i' Z_i' I_i' + v_i' \\ -\frac{dI_i'}{dr} &= j\hat{\kappa}_i' Y_i' V_i' + i_i' \end{aligned} \tag{2.47}$$

where

$$v_i'(r) = \hat{\kappa}_i' \iint M'(r, \theta, \varphi)\phi_i(\theta, \varphi)\, d\Omega - \frac{\hat{\kappa}_i'}{jk\epsilon} \iint J_r(r, \theta, \varphi)\phi_i(\theta, \varphi)\, d\Omega$$

$$i_i'(r) = \hat{\kappa}_i' \iint J'(r, \theta, \varphi)\phi_i(\theta, \varphi)\, d\Omega$$

$$\kappa_i' = \left(k^2 \epsilon\mu - \frac{\hat{\kappa}_i'^2}{r^2} \right)^{1/2}$$

$$Z_i' = \frac{1}{Y_i'} = \frac{\kappa_i'}{k\epsilon} \cdot$$

The partial differential equations (2.38b) can be reduced to a set of ordinary differential equations in a manner similar to the above. In this case one employs the representations

---

[6] $\partial/\partial\nu = \boldsymbol{\nu} \cdot {}_t\nabla$

$$V''(r, \theta, \varphi) = \sum_i V_i''(r) \frac{\psi_i(\theta, \varphi)}{\hat{\kappa}_i''}$$

(2.48)

$$I''(r, \theta, \varphi) = \sum_i I_i''(r) \frac{\psi_i(\theta, \varphi)}{\hat{\kappa}_i''}$$

where the characteristic modes ψ_i and the characteristic values $\hat{\kappa}_i''$ are defined by the eigenvalue problem

$$(r^2 \nabla_t \cdot {}_t\nabla + \hat{\kappa}_i''^2)\psi_i = 0 \tag{2.49}$$

with ψ_i subject to suitable boundary conditions. The orthonormality properties of ψ_i on the spherical cross-section,

$$\iint \psi_i \psi_j \, d\Omega = \delta_{ij} , \tag{2.50}$$

are readily deduced with the consequence that in (2.48)

$$V_i''(r) = \hat{\kappa}_i'' \iint V''(r, \theta, \varphi)\psi_i(\theta, \varphi) \, d\Omega$$

(2.51)

$$I_i''(r) = \hat{\kappa}_i'' \iint I''(r, \theta, \varphi)\psi_i(\theta, \varphi) \, d\Omega.$$

The transformation of equations (2.38b) in accord with the operations indicated in (2.51) then yields on use of the implied boundary conditions on ψ_i ,

$$\left[\psi_i \frac{\partial V''}{\partial \nu} - V'' \frac{\partial \psi_i}{\partial \nu}\right] = 0 \qquad \text{on } s, \tag{2.52}$$

the defining equations for the H-mode amplitudes:

$$-\frac{dV_i''}{dr} = j\kappa_i'' Z_i'' I_i'' + v_i''$$

(2.53)

$$-\frac{dI_i''}{dr} = j\kappa_i'' Y_i'' V_i'' + i_i''$$

where

$$v_i''(r) = \hat{\kappa}_i'' \iint M''(r, \theta, \varphi)\psi_i(\theta, \varphi) \, d\Omega$$

$$i_i''(r) = \hat{\kappa}_i'' \iint J''(r, \theta, \varphi)\psi_i(\theta, \varphi) \, d\Omega - \frac{\hat{\kappa}_i''}{jk\mu} \iint J_r(r, \theta, \varphi)\psi_i(\theta, \varphi) \, d\Omega$$

$$\kappa_i'' = (k^2 \epsilon\mu - \hat{\kappa}_i''^2/r^2)^{1/2}$$

$$Z_i'' = \frac{1}{Y_i''} = \frac{k\mu}{\kappa_i''} \cdot$$

Equations (2.47) and (2.53) are spherical transmission line equations of a more general form than the similar equations (2.13a) encountered in the φ independent analysis of Section $2a_1$. The chief difference between the two cases lies in the greater complexity of the source voltage v_i and source current i_i in the former case of arbitrary excitation.

Although the transmission line equations have been obtained in this section by a scalar representation of the potentials I', V'', etc., it is perhaps of interest to sketch the derivation of equations (2.47) and (2.53) by direct vector representation of the fields. In this case one starts with the transverse vector representations

$$
\begin{aligned}
r\mathbf{E}_t(r, \theta, \varphi) &= \sum_i \left[V_i'(r)\mathbf{e}_i'(\theta, \varphi) + V_i''(r)\mathbf{e}_i''(\theta, \varphi)\right] \\
r\mathbf{H}_t(r, \theta, \varphi) &= \sum_i \left[I_i'(r)\mathbf{h}_i'(\theta, \varphi) + I_i''(r)\mathbf{h}_i''(\theta, \varphi)\right]
\end{aligned}
\tag{2.54}
$$

where the single and double prime superscripts denote, respectively, the separation into E and H vector mode functions; this separation implies a restriction on the boundary conditions permissible on the cross-section of the spherical region under consideration. The vector mode functions in (2.54) are to be so chosen as to permit a simple evaluation of the mode amplitudes. Although the desired functions are evidently characteristic of the operator ${}_t\nabla\ \nabla_t$ in equations (2.35), we shall omit for brevity their detailed derivation and merely define them in terms of the already defined scalar mode functions $\phi_i(\theta, \varphi)$ and $\psi_i(\theta, \varphi)$. For the spherical regions in question the E-mode functions are defined by

$$
\begin{aligned}
-\hat{\kappa}_i'\mathbf{e}_i' &= r_t\nabla\phi_i \\
\hat{\kappa}_i'\phi_i &= r\nabla_t\cdot e_i'
\end{aligned}
\quad , \qquad \mathbf{h}_i' = \mathbf{r}_0 \times \mathbf{e}_i'
\tag{2.55a}
$$

and the H-mode functions by

$$
\begin{aligned}
-\hat{\kappa}_i''\mathbf{h}_i'' &= r_t\nabla\psi_i \\
\hat{\kappa}_i''\psi_i &= r\nabla_t\cdot\mathbf{h}_i''
\end{aligned}
\quad , \qquad \mathbf{e}_i'' = \mathbf{h}_i'' \times \mathbf{r}_0 \, .
\tag{2.55b}
$$

It is evident that the detailed equations (2.55) are equivalent to the scalar eigenvalue equations (2.42) and (2.49) and are generalizations of the corresponding φ independent equations (2.10a) and (2.10b). Moreover, from (2.55) one obtains on elimination of ϕ_i and ψ_i, respectively

$$
r^2{}_t\nabla\nabla_t\cdot\mathbf{e}_i' + \hat{\kappa}_i'^2\,\mathbf{e}_i' = 0 \tag{2.56a}
$$

$$
r^2{}_t\nabla\nabla_t\cdot h_i'' + \hat{\kappa}_i''^2 h_i'' = 0 \tag{2.56b}
$$

which, together with appropriate boundary conditions, constitute the vector eigenvalue problems for the E and H-modes, respectively. From equations

(2.56), or from (2.55) and the previously mentioned orthonormality properties of ϕ_i and ψ_i, one readily deduces

$$\iint \mathbf{e}_i' \cdot \mathbf{e}_j' \, d\Omega = \delta_{ij} = \iint \mathbf{e}_i'' \cdot \mathbf{e}_j'' \, d\Omega$$

(2.57)

$$\iint \mathbf{e}_i' \cdot \mathbf{e}_j'' \, d\Omega = 0$$

together with corresponding orthonormality properties for the $\mathbf{h}_i$; as above the surface integrals are to be extended over the entire cross-section. From (2.54) and (2.57) one then finds that

$$V_i = \iint r\mathbf{E}_t \cdot \mathbf{e}_i \, d\Omega$$

(2.58)

$$I_i = \iint r\mathbf{H}_t \cdot \mathbf{h}_i \, d\Omega$$

for both mode types. Transformation of equations (2.35) according to (2.58) then yields, after vector integration by parts, the same spherical transmission line equations (2.47) and (2.53) for the E and H mode amplitudes of (2.54). In terms of the vector mode functions $\mathbf{e}_i$, $\mathbf{h}_i$ the source terms are given by

$$v_i(r) = \iint r\mathbf{M} \cdot \mathbf{h}_i \, d\Omega - Z_i \iint r\,\mathbf{J} \cdot \mathbf{e}_{ri} \, d\Omega$$

(2.59)

$$i_i(r) = \iint r\,\mathbf{J} \cdot \mathbf{e}_i \, d\Omega - Y_i \iint r\mathbf{M} \cdot \mathbf{h}_{ri} \, d\Omega$$

where the r component vector functions $\mathbf{e}_{ri}$ and $\mathbf{h}_{ri}$ are defined by

$$\hat{\kappa}_i' \phi_i \mathbf{r}_0 = j\kappa_i' \mathbf{e}_{ri} , \qquad \hat{\kappa}_i'' \psi_i \mathbf{r}_0 = j\kappa_i'' \mathbf{h}_{ri} .$$

Equations (2.59) could equally well have been obtained from the corresponding expressions in (2.47) and (2.53) on use of (2.55); this involves an integration by parts assuming $M' = 0 = J'$ on the cross-sectional boundary. The modal superscripts have been omitted in equations (2.59) since the equations have the same form for both mode types providing one notes that for E-modes $\mathbf{h}_{ri} = 0$ whereas for H-modes $\mathbf{e}_{ri} = 0$. It should also be noted that the expressions for v_i and i_i in (2.59) do not require the explicit decomposition of $\mathbf{J}$ and $\mathbf{M}$ into their E and H-components.

3a. *Characteristic Problems in One Dimension*

As discussed in Section 2, the solution of a vector field problem, or of equivalent scalar problems, in the form of a representation requires the solution of both eigenvalue and transmission line problems. Although eigenvalue prob-

lems are in general multidimensional, their solution frequently may be traced back to corresponding one dimensional problems of the type considered in this section. From an operational point of view the one dimensional problems of interest are characteristic, in general, of a non-Hermitean Sturm-Liouville operator

$$L = -\frac{d}{dx}p(x)\frac{d}{dx} + q(x), \tag{3.1}$$

where p and q are assumed to be piecewise continuous in the interval $x_1 < x < x_2$. Associated with the operator L is a characteristic Green's function defined in the indicated interval by the inhomogeneous differential equation

$$[L - \lambda w(x)]G(x, x') = \delta(x - x') \tag{3.2a}$$

and subject to the boundary conditions

$$\left[p\frac{d}{dx} + \alpha_{1,2}\right]G(x, x') \to 0 \qquad \text{as } x \to x_{1,2}\,. \tag{3.2b}$$

The arbitrary complex parameter λ is to be so restricted as to insure the uniqueness of $G(x, x')$; the weight function $w(x)$ is a piecewise continuous function; and the delta function source term is defined by

$$\delta(x - x') = 0 \qquad \text{if } x \neq x', \qquad \int \delta(x - x')\,dx = 1,$$

the interval of integration including the singular point x'.

Also associated with the operator L are the set of characteristic functions $\phi_i(x)$ defined in the interval $x_1 < x < x_2$ by the homogeneous differential equations

$$[L - \lambda_i w(x)]\phi_i(x) = 0. \tag{3.3a}$$

The characteristic values λ_i are determined by subjecting $\phi_i(x)$ to boundary conditions defined in terms of the corresponding conditions (3.2b) on $G(x, x')$ by

$$\left[p\left(\phi_i\frac{dG}{dx} - G\frac{d\phi_i}{dx}\right)\right]_{x_1}^{x_2} \to 0 \tag{3.3b}$$

where the left hand side denotes the limiting difference of the bracketed quantity at $x \to x_2$ and $x \to x_1$. The boundary conditions on ϕ_i are phrased in the form (3.2b) to include the complete set of both proper (discrete) and improper (continuous) eigenfunctions; the former are square integrable in the given interval, the latter are not. For the proper eigenfunctions the boundary conditions (3.3b) may be reduced to those of the form (3.2b); this reduction is not possible for the improper functions. In view of the existence of finite solutions to equations (3.3a) there is a manifest ambiguity in the $G(x, x')$ defined by (3.2a) when $\lambda = \lambda_i$. Thus the uniqueness restriction on λ alluded to above is that $\lambda \neq \lambda_i$. For real p, q, w, and $\alpha_{1,2}$ (the Hermitean case) it can be shown that the eigenvalues λ_i of the operator L are real; hence for this case the restriction

$\lambda \neq \lambda$, reduces to $\mathscr{I}m\ \lambda \neq 0$ (provided the λ plane can be regarded as a simple surface).

By comparison with equations (2.14) and (2.23) it is seen that equations (3.2) characterize a transmission line problem in which $G(x, x')$ represents either the voltage V or the current I produced by a δ function source at $x = x'$. Although phrased as a second order differential equation problem, equations (3.2) could equally well have been put in the first order form more customary in general transmission line theory and more convenient when more general types of sources are treated. It is likewise evident that the characteristic function problem of (3.3) is a generalization of corresponding problems encountered in equations (2.11) and (2.17).

As previously stated and as also implied in (3.2) and (3.3) there exists an intimate connection between the characteristic Green's function $G(x, x')$ and the characteristic functions $\phi_i(x)$. As a consequence, the knowledge of $G(x, x')$ implies that of the $\phi_i(x)$, and conversely. This connection has been exploited in various ways by Weyl, Titchmarsh (loc. cit.) et al., to obtain the spectral representations (and hence the characteristic orthonormal functions) for a large number of operators L. In the following we shall employ essentially the same reasoning as the above authors to determine explicitly the orthonormal functions characteristic of the operators involved in the representation of fields in the spherical regions discussed in Section 2. However, the procedure to be employed below will involve the δ-function technique, whose oft discussed justification will not be considered herein. The virtue of this technique, besides that of simplicity, is that questions of "completeness" are answered naturally, albeit formally, by recourse to the basic concept of the δ function. For example, the existence of a complete set of orthonormal functions $\phi_i(x)$ presumes the ability to represent completely an arbitrary continuous[7] function in terms of the $\phi_i(x)$. If this arbitrary function is a δ function, the completeness and orthonormality of the $\phi_i(x)$ with respect to the weight function $w(x)$ is contained in the existence of the representation

$$\text{(3.4)} \qquad \frac{\delta(x - x')}{w(x')} = \sum_i \phi_i(x)\phi_i(x'), \qquad x_1 < x < x_2$$ [8]

the summation sign denoting here and in the following either or both a sum over a discrete index i characteristic of a discrete spectrum or an integral over a continuous index i characteristic of a continuous spectrum. That equation (3.4) is a completeness statement for an arbitrary continuous function $f(x)$ follows formally by multiplication of (3.4) by $f(x')w(x')$ and integration over x' from x_1 to x_2 , etc.; orthonormality follows formally by multiplication of (3.4) by

[7]Discontinuous function representations can likewise be considered by a simple "arithmetical mean" extension of the following considerations.

[8]Other non symmetrical forms of the completeness statement appropriate to Hermitean orthogonal and to bi-orthogonal representations will also be employed. In the former case $\phi_i(x')$ in (3.4) is to be replaced by $\phi_i^*(x')$, in the latter by $\psi_i(x')$ (an adjoint function).

$\phi_i(x')w(x')$ and integration from x_1 to x_2, etc. It should be noted that the representation statement for a function $f(x)$ can be cast in the form of the transform theorem

$$f(x) = \sum_i F_i \phi_i(x) \tag{3.4a}$$

$$F_i = \int_{x_1}^{x_2} f(x)\phi_i(x)\, dx \tag{3.4b}$$

where the sum sign is to be regarded in the general sense stated above. For a bi-orthogonal representation ϕ_i in (3.4b) is to be replaced by ψ_i.

Assuming the existence of the representation (3.4), one can investigate the connection between $G(x, x')$ and the $\phi_i(x)$. For by equation (3.4) the characteristic Green's function can be represented as

$$G(x, x') = \sum_i G_i \phi_i(x) \tag{3.5a}$$

$$\text{with} \qquad G_i = \int_{x_1}^{x_2} G(x, x')\phi_i(x)w(x)\, dx. \tag{3.5b}$$

The amplitudes G_i may be determined from (3.2a), the defining equation for $G(x, x')$, on multiplication by $\phi_i(x)$ and integration in accordance with (3.5b). On use of the self adjointness relation (Green's theorem) for the operator L

$$\int_{x_1}^{x_2} [\phi_i(x)LG(x, x') - G(x, x')L\phi_i(x)]\, dx = -\left[p\left(\phi_i \frac{dG}{dx} - G\frac{d\phi_i}{dx}\right)\right]_{x_1}^{x_2},$$

and of the defining equations (3.3) the $\phi_i(x)$, there is then obtained for[9]

$$G_i = -\frac{\phi_i(x')}{\lambda - \lambda_i} \tag{3.6}$$

whence

$$G(x, x') = -\sum_i \frac{\phi_i(x)\phi_i(x)}{\lambda - \lambda_i}. \tag{3.7}$$

The representation (3.7) obtained on the assumption of the completeness of the representation (3.4), indicates that $G(x, x')$ possesses singularities in the complex λ plane at the points λ_i. These singularities take the form of either poles or branch cuts depending on whether the λ_i characterize points of the discrete or continuous spectrum, respectively. In a purely formal way one can integrate (3.7) about a contour in the λ plane enclosing *all* the singularities of $G(x, x')$ and obtain by Cauchy's theorem the basic relation

$$-\frac{1}{2\pi i}\oint G(x, x')\, d\lambda = \sum_i \phi_i(x)\phi_i(x') = \frac{\delta(x - x')}{w(x')}. \tag{3.8}$$

[9]On use of (3.5b) in (3.6) one obtains a homogeneous integral equation for ϕ_i and λ_i which can be employed in a formulation of the eigenvalue problem alternative to that in equations (3.3).

The contour is taken around not only the poles of $G(x, x')$ but also the branch cuts if they exist. The rigorous justification of the above procedure has only been sketched. For Hermitean operators L a rigorous proof is contained in the works of Weyl, Titchmarsh, etc.[10] alluded to above. As is implied in the above sketch, difficulties in a rigorous proof arise when the operator L admits a continuous spectrum. It is then usually necessary to employ a limit process starting from the readily handled case wherein the only singularities of G in the λ plane are simple poles and then pass to the case wherein some or all of the poles coalesce into a branch cut or cuts characteristic of the continuous spectrum.

The contour integral relation (3.8) is the basis of a well defined procedure for the solution of the eigenvalue problem associated with an operator L. The virtue of this procedure is that the problem of finding a complete orthonormal set is reduced to that of constructing the characteristic Green's function $G(x, x')$ and completely investigating its singularities. Questions of uniqueness and dependence on boundary conditions of the characteristic orthonormal set are reduced to corresponding, and more easily treated, questions for a single characteristic Green's function satisfying a well defined inhomogeneous differential equation.

3b. *The Characteristic Green's Function (Resultant)*

The explicit construction of characteristic Green's functions involves the solution of one dimensional transmission line problems similar to those posed in Section 2. Equations (2.13a), for example, exhibit such problems in the general form of two simultaneous first order equations subject to boundary conditions. For *arbitrary* source distributions v and i it is frequently convenient to reduce the first order equations to second order equations of the form shown in (2.14a) or (2.14b) by means of a superposition argument, assuming first $i = 0$ and then $v = 0$. The latter equations are characterized by the operator L of (3.1) and hence their solution may be reduced, again utilizing the superposition argument, to the solution of the general Green's function[11] problem stated in equations (3.2).

For illustration, if in (3.2) with $x = r$ and

$$p(x) = \frac{1}{\epsilon(r)}, \qquad q(x) = -k^2\mu(r), \qquad \lambda = -\hat{\kappa}_i^2, \qquad w(x) = \frac{1}{\epsilon(r)r^2}, \tag{3.9a}$$

the Green's function solution is designated as $G_I(r, r')$, the solution of (2.14a) is given by superposition as

$$I_i(r) = jk \int G_I(r, r')v_i(r')\, dr', \tag{3.9b}$$

---

[10]loc. cit. Also cf. K. O. Friedrichs, "Spectral Representations of Linear Operators", lecture notes, New York University, 1948.

[11]If λ in (3.2) is an arbitrary complex parameter, $G(x, x')$ is designated as a "characteristic" Green's function (i.e. inverse operator $(L - \lambda)^{-1}$); if λ is fixed the adjective "characteristic" is omitted.

where the irrelevant prime superscript and the limits of integration have been omitted. Correspondingly if in (3.2) with $x = r$ and

$$p(x) = \frac{1}{\mu(r)}, \qquad q(x) = -k^2\epsilon(r), \qquad \lambda = -\hat{\kappa}_i^2, \qquad w(x) = \frac{1}{\mu(r)r^2}, \tag{3.10a}$$

the Green's function is designated as $G_V(r, r')$, the solution of (2.14b) (omitting the double prime superscript, etc.) is given by

$$V_i(r) = jk \int G_V(r, r')i_i(r')\, dr'. \tag{3.10b}$$

It has been tacitly assumed that the boundary conditions (3.2b) on $G_I(r, r')$ correspond to those on I_i whereas the conditions on $G_V(r, r')$ correspond to those on V_i ; for open regions wherein both V_i and I_i satisfy the same "radiation" conditions, $G_I = G_V$. Since (3.9b) is the solution of (2.13a) with $i = 0$ and since (3.10b) is the solution with $v = 0$, it follows by appropriate superposition that the general solution of (2.13a) is

$$\begin{aligned} I_i(r) &= jk \int G_I(r, r')v_i(r')\, dr' - \frac{1}{j\kappa_i Z_i}\frac{d}{dr}\int G_V(r, r')i_i(r')\, dr' \\ V_i(r) &= jk \int G_V(r, r')i_i(r')\, dr' - \frac{1}{j\kappa_i Y_i}\frac{d}{dr}\int G_I(r, r')v_i(r')\, dr'. \end{aligned} \tag{3.11}$$

This explicit solution is manifestly dependent on the solution of a Green's function problem of the type shown in (3.2).

Let us therefore briefly review the method of solution of the inhomogeneous differential equation (3.2a) subject to the boundary conditions (3.2b). One readily deduces by integration of (3.2a) about an infinitesimal interval centered at x', that the presence of the function source is equivalent to the demands that $p(dG/dx)$ possess at $x = x'$ a jump discontinuity of value -1 and that G be continuous at $x = x'$. At all other points $x \neq x'$ one likewise deduces that G satisfies the homogeneous form of (3.2a) and that both $p(dG/dx)$ and G are continuous; the latter property is of importance since p, q, and w are permitted to have jump discontinuities. A further interesting property of $G(x, x')$ readily derivable from (3.2) is that $G(x, x') = G(x', x)$ if p, q, w and $\alpha_{1,2}$ are real or complex. This symmetry property facilitates the explicit evaluation of $G(x, x')$. One considers two independent solutions of the homogeneous Sturm-Liouville Equation, the first obeying the boundary condition (3.2b) on G at $x = x_1$ and the second that on G at $x = x_2$. The two solutions are defined by

$$[L - \lambda w(x)]T(x) = 0, \qquad \left(p\frac{dT}{dx} + \alpha_1 T\right) = 0 \qquad \text{at } x = x_1 \tag{3.12a}$$

$$[L - \lambda w(x)]U(x) = 0, \qquad \left(p\frac{dU}{dx} + \alpha_2 U\right) = 0 \qquad \text{at } x = x_2 . \tag{3.12b}$$

In virtue of the above stated symmetry and continuity properties the characteristic Green's function of (3.2) can be expressed in terms of these solutions as

$$G(x, x') = \begin{cases} \dfrac{T(x)U(x')}{W(U, T)} & x < x' \\[2ex] \dfrac{T(x')U(x)}{W(U, T)} & x > x' \end{cases} \tag{3.13}$$

where $W(U, T)$ the Wronskian expression defined by

$$W(U, T) = p\left(U \frac{dT}{dx} - T \frac{dU}{dx}\right), \tag{3.14}$$

is independent of x in those regions wherein p, q, w and their derivatives are continuous. Since (3.13) satisfies (3.2a) and the boundary conditions (3.2b) and also possesses the required continuity properties at $x = x'$, it is apparent that it is the desired solution. The discontinuous representation in (3.13) is more succinctly stated in the form

$$G(x, x') = \frac{T(x_<)U(x_>)}{W(U, T)} \tag{3.15}$$

where the notation $x_<$ is employed for either x or x' depending on which is the smaller, and conversely for $x_>$. An even more compact notation is obtained if in addition T and U are so normalized that the Wronskian (3.14) is unity.

The constancy with x of the Wronskian can be employed to obtain an alternative expression for $G(x, x')$ that frequently facilitates the explicit construction of $G(x, x')$. Thus, if the Wronskian is evaluated at a convenient point, say x_0, and (3.15) is divided by $T(x_0)U(x_0)$, there is obtained

$$G(x, x') = \frac{t(x_<, x_0)\, u(x_>, x_0)}{-X(\lambda)} \tag{3.16}$$

where

$$t(x, x_0) = \frac{T(x)}{T(x_0)} \qquad u(x, x_0) = \frac{U(x)}{U(x_0)}$$

$$X^-(\lambda) = -\left(\frac{p}{T}\frac{dT}{dx}\right)_{x_0} \qquad X^+(\lambda) = \left(\frac{p}{U}\frac{dU}{dx}\right)_{x_0}$$

$$X = X^- + X^+.$$

General transmission line theory provides a ready physical interpretation of (3.16). As evident from equation (3.9), $G(x, x')$ can represent, for example, the current wave set up at any point x by a voltage point source of amplitude $1/jk$ at x'. Thus $t(x, x_0)$ and $u(x, x_0)$ are wave solutions normalized to unity at $x = x_0$, the former satisfying the prescribed terminal (impedance) conditions

at $x = x_1$, the latter at $x = x_2$. The quantity $X(\lambda)$ represents the total reactance ($= iZ(\lambda)$ where $Z(\lambda)$ is the total impedance) at x_0 and is the sum of the reactance $X^-(\lambda)$ and $X^+(\lambda)$ looking from x_0 in the negative and positive x directions, respectively. Alternatively, as in equation (3.10), $G(x, x')$ can represent the voltage wave set up by a point current source; in this case the denominator of (3.16) is denoted by $B(\lambda)$, the total susceptance ($= +iY(\lambda)$ where $Y(\lambda)$ is the total admittance) at x_0 . The simple zeros of the total reactance (or susceptance) define the so called resonances and resonant wave solutions of the transmission system; the corresponding λ values and distribution functions $t(x, x_0)$ are the eigenvalues and eigenfunctions of the discrete spectrum. One virtue of introducing impedance terminology in this connection is that the presence of a discrete spectrum, i.e. of resonances, have well established and intuitive answers in impedance theory. Moreover, the existence of systematic methods for expressing the reactance at any point say x_0 , in terms of the prescribed terminal reactances $\alpha_{1,2}$ provides a standardized formalism for constructing the characteristic Green's function of (3.16) almost at once. Although admittedly a question of terminology, impedance phraseology clothes a systematic procedure for obtaining solutions of arbitrary second order differential equations.

Let us return from the above digression to the explicit evaluation of the logarithmic derivatives X^- and X^+ together with the homogeneous solution $t(x, x_0)$ and $u(x, x_0)$ of equation (3.16). These can be expressed in terms of a set of regular (standing wave) fundamental solutions

$$c(x, x_0) \qquad \text{and} \qquad s(x, x_0)$$

satisfying the homogeneous equations in (3.12) and the real boundary conditions

$$\begin{aligned} c(x_0, x_0) &= 1 & \qquad s(x_0, x_0) &= 0 \\ p(x_0)c'(x_0, x_0) &= 0 & \qquad p(x_0)s'(x_0, x_0) &= 1 \end{aligned} \tag{3.17}$$

where the prime denotes a derivative with respect to the first argument. It follows from (3.12) that

$$\begin{aligned} t(x, x_0) &= c(x, x_0) - X^- s(x, x_0) \\ u(x, x_0) &= c(x, x_0) + X^+ s(x, x_0) \end{aligned} \tag{3.18}$$

where

$$X^- = \frac{p(x_1)\, c'(x_1, x_0) + \alpha_1\, c(x_1, x_0)}{p(x_1)\, s'(x_1, x_0) + \alpha_1\, s(x_1, x_0)} \tag{3.18a}$$

$$-X^+ = \frac{p(x_2)\, c'(x_2, x_0) + \alpha_2\, c(x_2, x_0)}{p(x_2)\, s'(x_2, x_0) + \alpha_2\, s(x_2, x_0)} . \tag{3.18b}$$

If $x_1(x_2)$ are singular points of the differential equation (3.2a), the "limit point"

or "limit circle" case is said to obtain at $x_1(x_2)$ when there is respectively an independence or regular dependence of $X^-(X^+)$ on the limiting terminal reactance $\alpha_1(\alpha_2)$. In terms of the solutions (3.18) the characteristic Green's function (3.16) may be written

$$G(x, x') = \frac{[c(x_<, x_0) - X^- s(x_<, x_0)][c(x_>, x_0) + X^+ s(x_>, x_0)]}{-X(\lambda)} \tag{3.19}$$

the singularities of which can be inferred from the regularity properties of X^- and X^+ in the λ plane. Three cases can be distinguished:

(1) *X^- and X^+ Meromorphic*

In this case the only singularities of G are simple poles (since $\partial X/\partial\lambda \neq 0$) located at the zeros λ_n of $X(\lambda)$. The completeness relation (3.8) yields on evaluation in (3.19) of the residues of $G(x, x')$ at $\lambda = \lambda_n$

$$\frac{\delta(x - x')}{w(x')} = \sum_n \frac{u_{\lambda_n}(x, x_0)\, u_{\lambda_n}(x', x_0)}{(\partial X/\partial\lambda)_{\lambda_n}}, \qquad X(\lambda_n) = 0. \tag{3.20}$$

The spectrum is evidently discrete. The eigenfunctions normalized to unity with respect to the weight function $w(x)$ are

$$\frac{u_{\lambda_n}(x, x_0)}{[(\partial X/\partial\lambda)_{\lambda_n}]^{1/2}} \equiv \frac{t_{\lambda_n}(x, x_0)}{[(\partial X/\partial\lambda)_{\lambda_n}]^{1/2}}\,. \tag{3.20a}$$

(2) *Only X^+ Meromorphic*

In this case the singularities of G take the form not only of simple poles located at the zeros λ_n of $X(\lambda)$ but also of branch cuts through the branch points of $X^-(\lambda)$. The completeness relation (3.8) yields after contour integration of (3.19)

$$\frac{\delta(x - x')}{w(x')} = \sum_n \frac{u_{\lambda_n}(x, x_0)\, u_{\lambda_n}(x', x_0)}{(\partial X/\partial\lambda)_{\lambda_n}} + \frac{1}{2\pi}\oint \frac{t(x_<, x_0)\, u(x_>, x_0)}{Z(\lambda)}\, d\lambda \tag{3.21a}$$

$$= \sum_n \frac{u_{\lambda_n}(x, x_0)\, u_{\lambda_n}(x', x_0)}{(\partial X/\partial\lambda)_{\lambda_n}} + \frac{1}{2\pi}\oint \frac{u(x, x_0)\, u(x', x_0)}{Z(\lambda)}\, d\lambda \tag{3.21b}$$

where $Z(\lambda) = -iX(\lambda)$. The sum represents the contribution of the residues at the poles of G; the contour integral (taken in the clockwise sense) represents the contribution from the branch cut. Equation (3.21a) represents the continuous eigenfunctions in biorthogonal form; the symmetrical representation of (3.21b) is obtained by noting that the X part of the $X^- = X - X^+$ term in $t(x, x_0)$ does not contribute to the branch cut integral. The normalized discrete and continuous eigenfunctions can be readily recognized in the latter representation.

(3) *X^-, X^+ not Meromorphic*

Both a discrete and a continuous spectrum are possible. The completeness statement (3.8) in this case can be written in the same biorthogonal form shown

in (3.21a). However, the symmetrical form[12] of the completeness relation becomes

$$\frac{\delta(x - x')}{w(x')} = \sum_n \frac{u_{\lambda_n}(x, x_0)\, u_{\lambda_n}(x', x_0)}{(\partial X/\partial\lambda)_{\lambda_n}}$$

$$+ \frac{1}{2\pi} \oint \left\{ \frac{c(x, x_0)c(x', x_0)}{Z(\lambda)} + \frac{s(x, x_0)s(x', x_0)}{Y(\lambda)} \right.$$

$$\left. + \frac{X^+(\lambda) - X^-(\lambda)}{2Z(\lambda)} [c(x, x_0)s(x', x_0) + c(x', x_0)s(x, x_0)] \right\} d\lambda \tag{3.22}$$

where $Z(\lambda) = -iX(\lambda)$ and $Y(\lambda) = -i(1/X^+ + 1/X^-)$. The discrete and continuous normalized eigenfunctions follow readily from the representation (3.22).

The above representations are illustrated in more detail in Sections 4 and 5 for some of the operators L encountered in spherical propagation problems. Representations of the form (3.20-2) apply even when p, q and w are discontinuous functions of x; the consequent discontinuous (in x) nature of the representations can be exhibited explicitly in such cases (cf. Sect. 4).

4. *One Dimensional Spectral Representations*

Several complete sets of eigenfunctions and eigenvalues characteristic of the one dimensional operators composing the spherical wave operator ∇^2 will be evaluated in this section. The operators in question are distinguished by the values of the parameters $p, q, w \cdots$ in the operator L defined in equation (3.1) and by the boundary conditions characterizing the domain of admissible functions. The spectral representations to be obtained will be utilized in Section 5.

$4a_1$. $p = \sin\theta$, $q = 0$, $w = \sin\theta$; $0 < \theta < \pi$

The characteristic Green's function (3.2a) is defined in the indicated domain by

$$\left[\frac{d}{d\theta}\sin\theta\frac{d}{d\theta} + \hat{\lambda}\sin\theta\right] G(\theta, \theta') = -\delta(\theta - \theta'). \tag{4.1}$$

The boundary points $\theta = 0, \pi$ are regular singular points of the differential operator and are of the "limit circle" type; however, rather than employ conditions of the form (3.2b), $G(\theta, \theta')$ is most simply characterized by conditions of finiteness at the boundaries $\theta = 0, \pi$. The Hermitean character of L and consequent reality of the eigenvalues implies that $G(\theta, \theta')$ is unique if $\mathscr{I}m\,\hat{\lambda} \neq 0$.

If $\hat{\lambda} = \nu(\nu + 1)$, solutions satisfying the homogeneous equations (3.12) and

[12] cf. Titchmarsh, loc. cit. Eq. (3.1.8).

the required finiteness conditions at $\theta = 0$ and π are, respectively, the Legendre functions

$$
\begin{aligned}
T(\theta) &= P_\nu(\cos\theta) \\
U(\theta) &= P_\nu(-\cos\theta).
\end{aligned}
\tag{4.2}
$$

The independence of these solutions for $\mathscr{I}m\,\nu \neq 0$ is assured by the non-vanishing of the Wronskian expression

$$
\sin\theta\left[P_\nu(\cos\theta)\frac{d}{d\theta}P_\nu(-\cos\theta) - P_\nu(-\cos\theta)\frac{d}{d\theta}P_\nu(\cos\theta)\right] = \frac{2}{\pi}\sin\nu\pi.
\tag{4.3}
$$

In accordance with equation (3.15) $G(\theta, \theta')$ is thus given by

$$
G(\theta, \theta') = -\frac{P_\nu(\cos\theta_<)P_\nu(-\cos\theta_>)}{(2/\pi)\sin\nu\pi}
\tag{4.4}
$$

whence by (3.8) the completeness relation in biorthogonal form follows as

$$
\frac{\delta(\theta - \theta')}{\sin\theta'} = \frac{1}{2\pi i}\oint\frac{P_\nu(\cos\theta_<)P_\nu(-\cos\theta_>)}{(2/\pi)\sin\nu\pi}\,d\hat{\lambda}
\tag{4.5}
$$

with the contour integral taken around all the singularities of the characteristic Green's function in the $\hat{\lambda}$ plane. The regularity of the Legendre functions and $\sin\nu\pi$ in the $\hat{\lambda}$ plane (note $P_\nu = P_{-\nu-1}$) implies that the only singularities are simple poles on the positive real $\hat{\lambda}$ axis at $\nu = n = 0, \pm1, \pm2 \cdots$, i.e. $\hat{\lambda} = (n + \frac{1}{2})^2 - \frac{1}{4}$. On evaluation of the integral (4.5) along the contour indicated

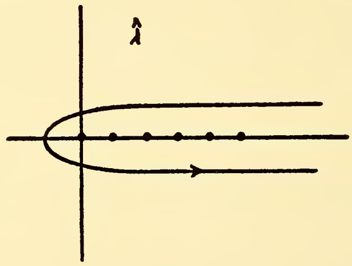


FIGURE 4.1

in Figure 4.1, one obtains by the Cauchy residue theorem as the symmetrical form of the completeness statement

$$
\begin{aligned}
\frac{\delta(\theta - \theta')}{\sin\theta'} &= \sum_{n=0}^{\infty}\frac{(n + \frac{1}{2})P_n(\cos\theta_<)P_n(-\cos\theta_>)}{(-1)^n} \\
&= \sum_{n=0}^{\infty}(n + \tfrac{1}{2})P_n(\cos\theta)P_n(\cos\theta')
\end{aligned}
\tag{4.6}
$$

since $P_n(x) = (-1)^n P_n(-x)$. By (3.8) it is evident that the normalized eigenfunctions are discrete and given by

$$\phi_i = (n + 1/2)^{1/2} P_n(\cos\theta), \tag{4.7}$$

the orthogonality being with respect to the weight function $\sin\theta$.

$4a_2$. $\theta_1 < \theta < \pi$

The characteristic Green's function is defined in the new domain by (4.1) but will be subject to the boundary conditions $(d/d\theta)\, G(\theta, \theta') = 0$ at the regular point $\theta = \theta_1$ and finiteness at the singular point $\theta = \pi$. As before $\mathcal{I}m\, \hat{\lambda} \neq 0$ with $\hat{\lambda} = \nu(\nu + 1)$.

The relevant solutions of the homogeneous equation are in this case

$$\begin{aligned} T(\theta) &= P_\nu(\cos\theta)\frac{d}{d\theta}P_\nu(-\cos\theta_1) - P_\nu(-\cos\theta)\frac{d}{d\theta}P_\nu(\cos\theta_1) \\ U(\theta) &= P_\nu(-\cos\theta) \end{aligned} \tag{4.8}$$

and hence by (3.16) with $x_0 = \theta_1$.

$$\begin{aligned} &G(\theta, \theta') \\ &= -\frac{\left[P_\nu(\cos\theta_<)\dfrac{d}{d\theta}P_\nu(-\cos\theta_1) - P_\nu(-\cos\theta_<)\dfrac{d}{d\theta}P_\nu(\cos\theta_1)\right]P_\nu(-\cos\theta_>)}{\left[\dfrac{2}{\pi}\sin\nu\pi\right]\left[\dfrac{d}{d\theta}P_\nu(-\cos\theta_1)\right]} \end{aligned} \tag{4.9}$$

The completeness relation is

$$\frac{\delta(\theta - \theta')}{\sin\theta'} = -\frac{1}{2\pi i}\oint G(\theta, \theta')\, d\hat{\lambda}$$

with the contour extended about all the singularities of G in the $\hat{\lambda}$ plane. The latter are in the form of simple poles occurring at the roots ν_i of

$$(d/d\theta)\, P_{\nu_i}(-\cos\theta_1) = 0 \tag{4.10}$$

which in view of the positive definite Hermitean character of L lie along the positive real $\hat{\lambda}$ axis. No poles lie at integral values of ν in virtue of the vanishing of the numerator of (4.9) at these points. Evaluation of the above integral along the same contour as in Figure 4.1 leads by the residue theorem to the discrete representation

$$\frac{\delta(\theta - \theta')}{\sin\theta'} = \sum_i \frac{(\nu_i + \frac{1}{2})P_{\nu_i}(-\cos\theta)P_{\nu_i}(-\cos\theta')}{\dfrac{\sin\nu_i\pi}{\pi}\left[\dfrac{d^2}{d\nu\, d\theta}P_\nu(-\cos\theta)\Big/\dfrac{d}{d\theta}P_\nu(\cos\theta)\right]_{\nu=\nu_i,\ \theta=}} \tag{4.11}$$

from which the normalized (with respect to the weight function $\sin \theta$) eigenfunctions are readily identified.

$4b_1$. $p = 1, q = -k^2, w = r^{-2}; r_1 < r < r_2$

The characteristic Green's function will be defined by

$$[(d^2/dr^2) + k^2 + (\lambda/r^2)]G(r, r') = -\delta(r - r') \tag{4.12}$$

with

$$k^2 \text{ a positive real constant, } \quad \text{and} \quad \frac{d}{dr} G = 0 \tag{4.13}$$

at the regular boundary points $r = r_1$ and r_2 . The operator L is thus Hermitean and hence the restriction $\mathscr{I}m\, \lambda \neq 0$ assures a unique solution to (4.12).

Solutions of the homogeneous operator equation possessing a vanishing derivative at r_1 and r_2 are if $\lambda = -\nu(\nu + 1)$

$$\begin{aligned} T(r) &= j_\nu(kr)n'_\nu(kr_1) - n_\nu(kr)j'_\nu(kr_1) \\ U(r) &= j_\nu(kr)n'_\nu(kr_2) - n_\nu(kr)j'_\nu(kr_2) \end{aligned} \tag{4.14}$$

where

$$j_\nu(x) = (\pi x/2)^{1/2} J_{\nu+\frac{1}{2}}(x) \quad \text{and} \quad n_\nu(x) = (\pi x/2)^{1/2} N_{\nu+\frac{1}{2}}(x) \tag{4.15}$$

are the spherical half order Bessel and Neumann functions[13] possessing a Wronskian $j_\nu(x)n'_\nu(d) - n_\nu(x)j'_\nu(x) = 1$. The Wronskian of the solutions (4.14) is

$$W_\nu(U, T) = k[j'_\nu(kr_1)n'_\nu(kr_2) - n'_\nu(kr_1)j'_\nu(kr_2)]$$

and hence by equation (3.15)

$$G(r, r') = \frac{[j_\nu(kr_<)n'_\nu(kr_1) - n_\nu(kr_<)j'_\nu(kr_1)][j_\nu(kr_>)n'_\nu(kr_2) - n_\nu(kr)j'_\nu(kr_2)]}{k[j'_\nu(kr_1)n'_\nu(kr_2) - n'_\nu(kr_1)j'_\nu(kr_2)]}$$

where, as above, $r_<$ and $r_>$ denote the lesser or greater, respectively, of the values r and r'. In view of the regularity of the functions T and U in the λ plane, the only singularities of the characteristic Green's function correspond to the zeros of the denominator in (4.16). The Hermitean and non-definite nature of the operator L imply that these singularities are located on both the positive and negative real λ axis and have the form of simple poles. In the vicinity of the typical pole $\lambda_i = -\nu_i(\nu_i + 1)$

$$W_\nu(U, T) \approx -\left[\frac{\partial}{\partial \nu} W_\nu\right]_{\nu_i} \frac{\lambda - \lambda_i}{2\nu_i + 1} + \cdots . \tag{4.17}$$

---

[13]This notation differs by a factor of x from the corresponding notation in Stratton, J. A., "Electromagnetic Theory", (1941), Sec. 7.4; the former appears most convenient in vector problems, the latter in scalar problems.

The completeness relation (3.8) is given by

$$r'^2\delta(r - r') = -\frac{1}{2\pi i}\oint G(r, r'; \lambda)\, d\lambda$$

with the contour as shown in Figure 4.2 extended about all the singularities of G.

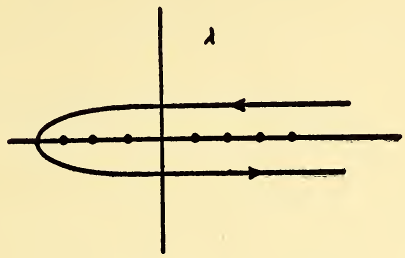


FIGURE 4.2

Evaluation of the residues at the poles then yields

$$r'^2\delta(r - r') = \sum_{\nu_i} (2\nu_i + 1) \tag{4.18}$$
$$\cdot \frac{[j_{\nu_i}(kr)n'_{\nu_i}(kr_1) - n_{\nu_i}(kr)j'_{\nu_i}(kr_1)][j_{\nu_i}(kr')n'_{\nu_i}(kr_1) - n_{\nu_i}(kr)j'_{\nu_i}(kr_1)]}{[(\partial/\partial\nu)W_\nu]_{\nu_i}}$$

from which the spectrum is manifestly discrete with eigenfunctions

$$\left\{\frac{2\nu_i + 1}{[(\partial/\partial\nu)W_\nu]_{\nu_i}}\right\}^{1/2} [j_{\nu_i}(kr)n'_{\nu_i}(kr_1) - n_{\nu_i}(kr)j'_{\nu_i}(kr_1)] \tag{4.19a}$$

normalized to unity with respect to the weight function r^{-2}, and with eigenvalues $\lambda_i = -\nu_i\,(\nu_i + 1)$ given by

$$W_{\nu_i}(U, T) = 0 \tag{4.19b}$$

4b$_2$. $0 < r < r_2$

The characteristic Green's function will be defined in the domain $0 < r < r_2$ as in equation (4.12). Let $\lambda = -\nu(\nu + 1)$ or more definitely $\nu + \frac{1}{2} = -i(\lambda - \frac{1}{4})^{1/2}$, and choose

$$\mathcal{R}e\,(\nu + \tfrac{1}{2}) = \mathcal{I}m\,(\lambda - \tfrac{1}{4})^{1/2} > 0. \tag{4.20}$$

The singular point at $r = 0$ is then of the "limit point" type and hence for a unique characteristic Green's function the boundary conditions may be given in the form

$$\begin{aligned} &G \text{ finite} && \text{at } r = 0 \\ &dG/dr = 0 && \text{at } r = r_2\,. \end{aligned} \tag{4.21}$$

The desired homogeneous solutions (3.12) are

$$U(r) = j_\nu(kr)n'_\nu(kr_2) - n_\nu(kr)j'_\nu(kr_2)T(r) = j_\nu(kr) \tag{4.22}$$

and possess a Wronskian

$$W_\nu(U, T) = kj'_\nu(kr_2).$$

The completeness relation (3.8) then follows from the resulting characteristic Green's function (3.15) as

$$r'^2\delta(r - r') = -\frac{1}{2\pi i}\oint \frac{j_\nu(kr_<)[j_\nu(kr_>)n'_\nu(kr_2) - n_\nu(kr)j'_\nu(kr_2)]}{kj'_\nu(kr_2)}\,d\lambda \tag{4.23}$$

with the contour extended about all the singularities of the integrand in the λ plane. Since the operator L is still Hermitean for this case, all singularities lie on the real λ axis. There may be a discrete number of singularities in the form of simple poles located on the real axis $\lambda < \frac{1}{4}$ at the zeros of $j'_\nu(kr_2)$, if any; and there is also a continuous singularity in the form of a branch cut extending along the real axis from $\lambda = \frac{1}{4}$ to $+\infty$. The latter delimits the positive imaginary branch of $(\lambda - \frac{1}{4})^{1/2}$ on which $j_\nu(kr)/j'_\nu(kr_2)$ is regular. The required contour for the integral in (4.23) is shown in Figure 4.3. On evaluating the residues at

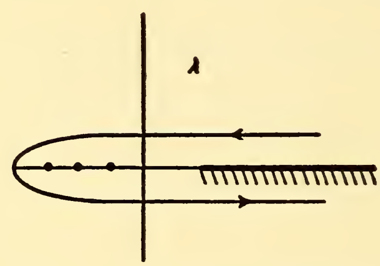


FIGURE 4.3

the poles, and expressing the branch cut integral in terms of ν, one obtains for equation (4.23)

$$\begin{aligned} r'^2\delta(r - r') &= \sum_{\nu_i}(2\nu_i + 1)\frac{j_{\nu_i}(kr)j_{\nu_i}(kr')}{k[(\partial/\partial\nu)j'_\nu(kr_2)/n'_\nu(kr_2)]_{\nu_i}} \\ &\quad + \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{j_\nu(kr)[j_\nu(kr')n'_\nu(kr_2) - n_\nu(kr')j'_\nu(kr_2)]}{kj'_\nu(kr_2)}(2\nu + 1)\,d\nu \end{aligned} \tag{4.24}$$

where $c = -\frac{1}{2} + 0$ and the ν_i are determined by

$$j'_{\nu_i}(kr_2) = 0. \tag{4.25}$$

It is to be noted that (4.25) has a finite number of solutions ν_i for $kr_2 > \pi$ but finite. In view of the symmetry in $\nu + \frac{1}{2}$ of the path of integration, the integral

over ν in (4.24) is symmetrical in r and r'.[14] From the representation in (4.24) it is evident that the spectrum possesses in general both a discrete and a continuous part. The non-diagonal (biorthogonal) integral operator in (4.24) can be readily cast into a diagonal form (cf. equations 3.21a and b) from which both the discrete and continuous eigenfunctions are directly obtained; however, the representation in (4.24) indicating the continuous eigenfunctions in biorthogonal form is frequently more useful.

$4b_3$. $r_1 < r < \infty$

A characteristic Green's function in the interval $r_1 < r < \infty$ will again be defined as in equation (4.12) but subject to the boundary conditions

$$\begin{aligned} &dG/dr = 0 && \text{at } r = r_1 \\ &dG/dr - ikG \to 0 && \text{at } r \to \infty. \end{aligned} \tag{4.26}$$

The latter so called "radiation" condition with k real may be phrased alternatively as the condition $G \to 0$ at $r \to \infty$ if $\mathscr{I}m\, k > 0$. In view of the complex nature of the latter condition it is evident that the operator is non-Hermitean and hence to assure a unique G, λ must be suitably restricted. For k positive real a simple evaluation of the location of the eigenvalues via Green's theorem shows that the desired restriction is $\mathscr{I}m\, \lambda > 0$. Moreover, it is of interest to note that although in the Hermitean case the singular point at infinity is of the "limit circle" type, for the non-Hermitean case with complex k it becomes a "limit point" case.

The homogeneous solutions satisfying the boundary conditions (4.26) are, respectively, if $\lambda = -\nu(\nu + 1)$

$$\begin{aligned} T(r) &= j_\nu(kr)n'_\nu(kr_1) - n_\nu(kr)j'_\nu(kr_1) \\ U(r) &= h_\nu^{(1)}(kr) \end{aligned} \tag{4.27}$$

where in accordance with the previous definitions (4.15)

$$h_\nu^{(1)}(x) = (\pi x/2)^{1/2} H_{\nu+\frac{1}{2}}^{(1)}(x)$$

defines the spherical Hankel function. The Wronskian of the two solutions (4.27) is

$$W_r(U, T) = -kh_\nu^{(1)\prime}(kr_1). \tag{4.28}$$

The completeness relation follows by (3.8) and (3.15) as

$$r'^2 \delta(r - r') = \frac{1}{2\pi i} \oint \frac{[j_\nu(kr_<)n'_\nu(kr_1) - n_\nu(kr)j'_\nu(kr_1)]h_\nu^{(1)}(kr_>)}{k\, h_\nu^{(1)\prime}(kr_1)}\, d\lambda \tag{4.29}$$

---

[14] Note that $n_\nu(x) = [\cos (\nu + \frac{1}{2})\pi\, j_\nu(x) - j_{-\nu}(x)]/\sin (\nu + \frac{1}{2})\pi$ where $x = kr'$.

the contour being extended about all the singularities of the integrand in the λ plane. Since $T(r)$ and $h_\nu^{(1)}(kr)/h_\nu^{(1)\prime}(kr_1)$ are integral functions of λ, the only singularities occur at the roots ν_i of

$$h_{\nu_i}^{(1)\prime}(kr_1) = 0 \tag{4.30}$$

and have the form of simple poles. The relevant roots are tabulated in Bremmer (loc. cit.) p. 44 and are found to occur in the lower half plane $\mathscr{I}m\ \lambda < 0$ or in the first quadrant of the ν-plane. The contour of integration for (4.29) is somewhat as shown in Figure 4.4

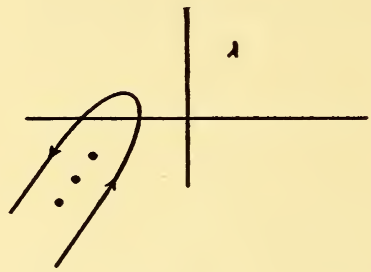


FIGURE 4.4

On evaluation of the residues at the indicated poles there is obtained for (4.29)

$$r'^2\delta(r - r') = \sum_{\nu_i} (2\nu_i + 1) \frac{h_{\nu_i}^{(1)}(kr)h_{\nu_i}^{(1)}(kr')}{ik[(\partial/\partial\nu)h_\nu^{(1)\prime}(kr_1)/j_\nu'(kr_1)]_{\nu_i}} \tag{4.31}$$

where the sum is to be taken over all the roots ν_i of (4.30). The spectrum is evidently discrete. The eigenfunctions normalized with respect to the weight function r^{-2} are

$$\left[\frac{2\nu_i + 1}{ik[(\partial/\partial\nu)h_\nu^{(1)\prime}(kr_1)/j_\nu'(kr_1)]_{\nu_i}}\right]^{1/2} h_{\nu_i}^{(1)}(kr). \tag{4.32}$$

It is to be emphasized that in view of the non-Hermitean nature of the operator L, orthogonality is to be understood in the symmetric and not in the Hermitean (complex conjugate) sense.

$4b_4$. $0 < r < \infty$

The characteristic Green's function is defined in equation (4.12). As in examples (b_2) and (b_3) let us choose $\mathscr{R}e\ (\nu + \frac{1}{2}) = \mathscr{I}m\ (\lambda - \frac{1}{4})^{1/2} > 0$ and $\mathscr{I}m\ k > 0$. It is thereby implied that the singular points $r = 0$ and ∞ are of the "limit point" type and hence requirements of finiteness at these points suffice to uniquely characterize the $G(r, r')$ of equation (4.12).

The required homogeneous solutions are

$$\begin{aligned} T(r) &= j_\nu(kr) \\ U(r) &= h_\nu^{(1)}(kr) \end{aligned} \tag{4.33}$$

whose Wronskian is $-ik$. By (3.15) the explicit characteristic Green's function becomes

$$G(r, r') = \frac{j_\nu(kr_<)h_\nu^{(1)}(kr_>)}{-ik}\,; \tag{4.34}$$

it possesses a branch point at $\lambda = \frac{1}{4}$ and is regular in that branch of the Riemann surface of $(\lambda - \frac{1}{4})^{1/2}$ for which $\mathscr{I}m\ (\lambda - \frac{1}{4})^{1/2} > 0$, i.e. excluding a branch cut along the real λ axis from $\lambda = \frac{1}{4}$ to ∞. On forming the contour integral of $G(r, r')$ about the branch cut (cf. Figure 4.5), one obtains by equation (3.8) the spectral representation

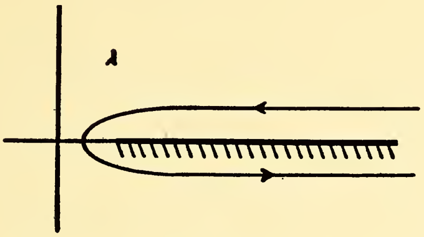


FIGURE 4.5

of the identity operator as

$$r'^2\delta(r - r') = -\frac{1}{2\pi}\oint \frac{j_\nu(kr_<)h_\nu^{(1)}(kr_<)}{k}\,d\lambda \tag{4.35a}$$

$$= \frac{1}{2\pi k}\int_{c-i\infty}^{c+i\infty} j_\nu(kr)h_\nu^{(1)}(kr')(2\nu + 1)\,d\nu\ [15] \tag{4.35b}$$

with $c = -\frac{1}{2} + 0$.

The spectrum for this case is manifestly continuous. The non-diagonal, or biorthogonal, form[16] of the completeness relation in (4.34) is usually most convenient as the basis for a transform theorem. It is of interest to point out the form of completeness relation deducible from the static (i.e. $k^2 \to 0$) characteristic Green's function. On substitution of the small argument asymptotic relations

$$\begin{aligned} j_\nu(x) &\sim \frac{\pi^{1/2}}{\Gamma(\nu + \frac{3}{2})}\left(\frac{x}{2}\right)^{\nu+1} \\ h_\nu^{(1)}(x) &\sim \frac{-i(\pi)^{1/2}}{\sin(\nu + \frac{1}{2})\pi\Gamma(\frac{1}{2} - \nu)}\left(\frac{x}{2}\right)^{-\nu} \end{aligned} \tag{4.36}$$

---

[15]cf. N. N. Lebedev, Dokl. Acad. Nauk. USSR, (1947), Vol. 58, No. 6.

[16]The symmetrical form with eigenfunctions proportional to $h_\nu^{(1)}(kr)$ is obtained as in the transition from equations (3.21a) to (3.21b).

and the identity

$$\cos \nu\pi = \Gamma(\tfrac{1}{2} + \nu)\Gamma(\tfrac{1}{2} - \nu)$$

into (4.35b), there is obtained in the limit $k^2 \to 0$

$$r'^2 \delta(r - r') = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} r^{\nu+1} r'^{\nu} \, d\nu \tag{4.37}$$

which is, of course, the basis for the Mellin transform.

4c$_1$. $p = \dfrac{1}{\epsilon(r)}$, $\quad q = -k^2, \quad w = \dfrac{1}{\epsilon(r)r^2}$; $\quad r_1 < r < r_3$

A characteristic Green's function is defined in the interval $r_1 < r < r_3$ by

$$\left(\frac{d}{dr}\frac{1}{\epsilon(r)}\frac{d}{dr} + k^2 + \frac{\lambda}{\epsilon(r)r^2}\right)G(r, r') = -\delta(r - r') \tag{4.38}$$

and subject to the boundary conditions $dG/dr = 0$ at the regular points $r = r_1$ and r_3. The parameter k^2 is a positive real constant and

$$\epsilon(r) \begin{array}{ll} = \epsilon_1 & r_1 < r < r_2 \\ = \epsilon_2 & r_2 < r < r_3 \end{array}$$

with ϵ_1 and ϵ_2 real constants. It follows from (4.38) that both $(1/\epsilon(r))/(dG/dr)$ and G are continuous at all points $r \neq r'$. In view of the reality of the operator L the restriction $\mathcal{I}m\, \lambda \neq 0$ assures a unique $G(r, r')$.

The discontinuous nature of $\epsilon(r)$ implies a like character for the solutions of the homogeneous form of equation (4.38). Thus if $\lambda = -\nu(\nu + 1)$, the solution $t(r, r_2)$ with vanishing derivative at $r = r_1$, with $(1/\epsilon)$ (dt/dr) and t continuous at $r = r_2$, and normalized to unity at $r = r_2$ is piecewise represented by

$$t(r, r_2) = c_\nu(k_\alpha r, k_\alpha r_2) - X_\nu^- \frac{s_\nu(k_\alpha r, k_\alpha r_2)}{k_\alpha/\epsilon_\alpha}. \tag{4.39}$$

The index α denotes 1 or 2 depending on whether $r_1 < r < r_2$ or $r_2 < r < r_3$, $k_\alpha = k(\epsilon_\alpha)^{1/2}$, and in accordance with the format in equations (3.17-8) and the definitions (4.15)

$$c_\nu(x, y) = j_\nu(x)n_\nu'(y) - n_\nu(x)j_\nu'(y)$$

$$s_\nu(x, y) = j_\nu(y)n_\nu(x) - n_\nu(y)j_\nu(x) \tag{4.39a}$$

$$X_\nu^- = \frac{k_1}{\epsilon_1}\frac{c_\nu'(k_1 r_1, k_1 r_2)}{s_\nu'(k_1 r_1, k_1 r_2)}.$$

Correspondingly, the solution $u(r, r_2)$, normalized to unity at $r = r_2$ and with vanishing derivative at $r = r_3$, is given by

$$u(r, r_2) = c_\nu(k_\alpha r, k_\alpha r_2) + X_\nu^+ \frac{s_\nu(k_\alpha r, k_\alpha r_2)}{k_\alpha/\epsilon_\alpha} \tag{4.40}$$

$$X_\nu^+ = -\frac{k_2}{\epsilon_2} \frac{c_\nu'(k_2 r_3, k_2 r_2)}{s_\nu'(k_2 r_3, k_2 r_2)} . \tag{4.40a}$$

In view of equations (4.39-40) the characteristic Green's function has the discontinuous representation (cf. 3.19)

$$G(r, r') = \frac{\left[c_\nu(k_\alpha r_<, k_\alpha r_2) - X_\nu^- \dfrac{s_\nu(k_\alpha r_<, k_\alpha r_2)}{k_\alpha/\epsilon_\alpha}\right]\left[c_\nu(k_\alpha r_>, k_\alpha r_2) + X_\nu^+ \dfrac{s_\nu(k_\alpha r_>, k_\alpha r_2)}{k_\alpha/\epsilon_\alpha}\right]}{-X_\nu} \tag{4.41}$$

where

$$X_\nu = X_\nu^- + X_\nu^+ . \tag{4.41a}$$

The meromorphic character of $t(r, r_2)$ and $u(r, r_2)$ implies that the only singularities of $G(r, r')$ in the λ plane lie at the zeros ν_i of X_ν. The Hermitean and non-definite character of the operator L imply that these singularities are located on both the positive and negative real λ axis and in fact have the form of simple poles. As in equation (3.20) the contour integral of the characteristic Green's function along a path like that shown in Figure 4.2 leads on evaluation of residues to the completeness relation

$$\frac{\delta(r - r')}{1/\epsilon(r')r'^2} = \sum_{\nu_i} \frac{(2\nu_i + 1)}{(\partial X_\nu/\partial\nu)_{\nu_i}} \left[c_{\nu_i}(k_\alpha r, k_\alpha r_2) - X_{\nu_i}^- \frac{s_{\nu_i}(k_\alpha r, k_\alpha r_2)}{k_\alpha/\epsilon_\alpha}\right] \cdot \left[c_{\nu_i}(k_\alpha r', k_\alpha r_2) - X_{\nu_i}^- \frac{s_{\nu_i}(k_\alpha r', k_\alpha r_2)}{k_\alpha/\epsilon_\alpha}\right] \tag{4.42}$$

The eigenfunctions normalized to unity with respect to a weight function $\epsilon(r)^{-1}r^{-2}$ thus have the discontinuous representation

$$\begin{aligned} &\left[\frac{2\nu_i + 1}{(\partial X_\nu/\partial\nu)_{\nu_i}}\right]^{1/2} \frac{c_{\nu_i}(k_1 r, k_1 r_1)}{c_{\nu_i}(k_1 r_2, k_1 r_1)}, \qquad r_1 < r < r_2 \\ &\left[\frac{2\nu_i + 1}{(\partial X_\nu/\partial\nu)_{\nu_i}}\right]^{1/2} \frac{c_{\nu_i}(k_2 r, k_2 r_3)}{c_{\nu_i}(k_2 r_2, k_2 r_3)}, \qquad r_2 < r < r_3 , \end{aligned} \tag{4.43}$$

with ν_i determined by

$$X_{\nu_i} = 0.$$

By comparison of (4.43) with the eigenfunctions (4.19a) of example b_1 it is

evident that the latter are just the special case $\epsilon_1 = \epsilon_2 = 1$ and $r_2 = r_3$ of the former in a somewhat different form.

$4c_2$. $0 < r < r_3$

The characteristic Green's function is defined as in (4.38) but with $r_1 = 0$. Since $r = 0$ is a singular point of the "limit point" variety (provided $\mathcal{R}e\,(\nu + \frac{1}{2}) > 0$), the results of the previous example c_1 are directly applicable on letting $r_1 \to 0$. The nature of the spectrum, however, becomes more apparent on direct investigation of the present case.

With the same choice of parameters as in c_1 but with $\mathcal{R}e\,(\nu + \frac{1}{2}) = \mathcal{I}m\,(\lambda - \frac{1}{4})^{1/2} > 0$ the desired homogeneous solutions $t(r, r_2)$ and $u(r, r_2)$, the former finite at $r = 0$ and the latter with vanishing derivative at $r = r_3$, are the same as in equations (4.39) and (4.40) with the sole modification

$$X_\nu^- = \frac{k_1}{\epsilon_1}\frac{j_\nu'(k_1 r_2)}{j_\nu(k_1 r_2)}, \tag{4.44}$$

which implies that for $r \leq r_2$

$$t(r, r_2) = \frac{j_\nu(k_1 r)}{j_\nu(k_1 r_2)}.$$

The characteristic Green's function possesses the same discontinuous representation given in (4.41) but with X_ν^- given by (4.44). In view of the branch point singularity of X_ν^- at $\lambda = \frac{1}{4}$ the solution $t(r, r_2)$ is no longer a meromorphic function of λ. Hence the singularities of the characteristic Green's function in the λ plane are not only simple poles located at the zeros of X_ν (cf. equation (4.41a)) but also a branch cut along the real λ axis from $\lambda = \frac{1}{4}$ to ∞. The associated spectrum of the operator L for this case thus comprises not only a discrete but also a continuous set of eigenvalues; it is the continuous set that is not explicitly evident on taking the limit $r_1 \to 0$ of the results in c_1.

The completeness relation (3.8) is obtained by integration of the characteristic Green's function along a contour in the λ plane similar to that shown in Figure 4.3. As in example b_2 the integral around the poles is expressed in terms of residues; however, the branch cut integral is expressed in terms of the variable ν. The completeness relation thus takes the form

$$\begin{aligned}\epsilon(r')r'^2\delta(r - r') = &\sum_{\nu_i}\frac{(2\nu_i + 1)}{(\partial X_\nu/\partial\nu)_{\nu_i}}\left[c_{\nu_i}(k_\alpha r, k_\alpha r_2) - X_{\nu_i}^-\frac{s_{\nu_i}(k_\alpha r, k_\alpha r_2)}{k_\alpha/\epsilon_\alpha}\right]\\ &\cdot\left[c_{\nu_i}(k_\alpha r', k_\alpha r_2) - X_{\nu_i}^-\frac{s_{\nu_i}(k_\alpha r', k_\alpha r_2)}{k_\alpha/\epsilon_\alpha}\right]\\ &+\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\left[c_\nu(k_\alpha r, k_\alpha r_2) - X_\nu^-\frac{s_\nu(k_\alpha r, k_\alpha r_2)}{k_\alpha/\epsilon_\alpha}\right]\\ &\cdot\left[c_\nu(k_\alpha r', k_\alpha r_2) + X_\nu^+\frac{s_\nu(k_\alpha r', k_\alpha r_2)}{k_\alpha/\epsilon_\alpha}\right]\frac{2\nu + 1}{X_\nu}\,d\nu\end{aligned} \tag{4.45}$$

with $c = -\frac{1}{2} + 0$. The symmetry in r and r' of the integrand in (4.45) is a consequence of the evenness in $\nu + \frac{1}{2}$ of the functions c_ν and s_ν (cf. equation 3.32). The discrete eigenfunctions, normalized to unity with respect to the weight function $1/\epsilon(r)r^2$, follow from (4.45) as

$$
(4.46) \qquad
\begin{aligned}
&\left[\frac{2\nu_i + 1}{(\partial X_\nu/\partial \nu)_{\nu_i}}\right]^{1/2} \frac{j_{\nu_i}(k_1 r)}{j_{\nu_i}(k_1 r_2)}, && 0 < r < r_2 \\
&\left[\frac{2\nu_i + 1}{(\partial X_\nu/\partial \nu)_{\nu_i}}\right]^{1/2} \frac{c_{\nu_i}(k_2 r, k_2 r_3)}{c_{\nu_i}(k_2 r_2, k_2 r_3)}, && r_2 < r < r_3
\end{aligned}
$$

with the ν_i determined by

$$
X_{\nu_i} = \frac{k_1}{\epsilon_1} \frac{j'_{\nu_i}(k_1 r_2)}{j_{\nu_i}(k_1 r_2)} - \frac{k_2}{\epsilon_2} \frac{c'_{\nu_i}(k_2 r_3, k_2 r_2)}{s'_{\nu_i}(k_2 r_3, k_2 r_2)} = 0.
$$

The continuous eigenfunctions are displayed in biorthogonal form in the ν integral of (4.45); the eigenfunctions in symmetrical form are proportional to $u(r, r_2)$ as is evident from the corresponding form in equation (3.21b). The above results should be compared with those for the special case $\epsilon_1 = \epsilon_2 = 1$ and $r_2 = r_3$ discussed in example b_2.

4c$_3$. $r_1 < r < \infty$

The characteristic Green's function will be defined by equation (4.38) but subject to the boundary conditions

$$
\frac{dG}{dr} = 0 \qquad \text{at } r = r_1
$$

$$
G \to 0 \qquad \text{as } r \to \infty .
$$

The parameter k^2 is now chosen as complex with a positive imaginary part (however small) and

$$
\epsilon(r) \begin{cases} = \epsilon_1 , & r_1 < r < r_2 \\ = \epsilon_2 , & r_2 < r. \end{cases}
$$

The operator L is non-Hermitean in this case and hence $\mathcal{I}m\ \lambda \neq 0$ does not suffice to insure a unique $G(r, r')$; rather, as in example b_3, it is necessary to restrict λ to $\mathcal{I}m\ \lambda > 0$. In view of the "limit point" character of the singular point at ∞ when k^2 is complex, it is not possible to consider this case as the limiting form $r_3 \to \infty$ of case c_1 in which, since k^2 is real, the point at infinity is of the "limit circle" type.

Solutions of the homogeneous equation (4.38) possessing the requisite continuity properties at $r = r_2$ and satisfying the boundary conditions at $r = r_1$

and ∞ are given, respectively, by equations (4.39) and (4.40) except that in this case

$$X_\nu^+ = \frac{k_2}{\epsilon_2}\frac{h_\nu^{(1)'}(k_2r_2)}{h_\nu^{(1)}(k_2r_2)} \tag{4.47}$$

which implies that for $r > r_2$

$$u(r, r_2) = \frac{h_\nu^{(1)}(k_2r)}{h_\nu^{(1)}(k_2r_2)}\,.$$

The characteristic Green's function is discontinuously represented as in equation (4.41). Since

$$H^{(1)}_{(\nu+\frac{1}{2})}(x) = \exp\{-i(\nu + \tfrac{1}{2})\pi\}H^{(1)}_{-(\nu+\frac{1}{2})}(x)$$

it follows that X_ν^+ is an even function of ν, and hence an integral or meromorphic function of λ, as is X_ν^-. Thus the only singularities of the characteristic Green's function are simple poles located in the λ plane at the zeros of X_ν. The location of these poles is for the most part that indicated in example b_3; however, in addition to the poles located in the fourth quadrant of the λ plane, there may occur poles on the real λ axis when $\epsilon_1/\epsilon_2 > 1$.

On integration of the characteristic Green's function about a contour (cf. Figure 4.4) enclosing all its singularities, one obtains the completeness relation (3.8) in the form given in equation (4.42). The spectrum is evidently discrete. The eigenfunctions normalized to unity with respect to a weight function $1/\epsilon(r)r^2$ become in this case

$$\begin{aligned}
&\left[\frac{2\nu_i + 1}{(\partial X_\nu/\partial\nu)_{\nu_i}}\right]^{1/2} \frac{c_{\nu_i}(k_1r,\ k_1r_1)}{c_{\nu_i}(k_1r_2\ ,\ k_1r_1)}\,, \qquad r_1 < r < r_2\\
&\left[\frac{2\nu_i + 1}{(\partial X_\nu/\partial\nu)_{\nu_i}}\right]^{1/2} \frac{h_{\nu_i}^{(1)}(k_2r)}{h_{\nu_i}^{(1)}(k_2r_2)}\,, \qquad r_2 < r
\end{aligned} \tag{4.48}$$

where the ν_i are determined from the roots of

$$X_\nu = \frac{k_1}{\epsilon_1}\frac{c_\nu'(k_1r_1\ ,\ k_1r_2)}{s_\nu'(k_1r_1\ ,\ k_1r_2)} + \frac{k_2}{\epsilon_2}\frac{h_\nu^{(1)'}(k_2r_2)}{h_\nu^{(1)}(k_2r_2)} = 0.$$

As in the special case $\epsilon_1 = \epsilon_2 = 1$, $r_1 = r_2$ treated in example b_3 orthogonality is to be understood in the symmetrical sense despite the complex nature of the eigenfunctions.

$4c_4$. $0 < r < \infty$

The characteristic Green's function is defined by equation (4.38) and subject to boundary conditions

$$G \text{ finite as } r \to 0$$

$$G \text{ zero as } r \to \infty$$

with the parameters k^2 and $\epsilon(r)$ defined as in the previous example c_3 except that $r_1 = 0$. The restrictions $\mathcal{R}e\ (\nu + \frac{1}{2}) = \mathcal{I}m\ (\lambda - \frac{1}{4})^{1/2} > 0$ and $\mathcal{I}m\ k^2 > 0$, respectively, insure that the singular points $r = 0$ and $r = \infty$ are of the "limit point" type. The further restriction $\mathcal{I}m\ \lambda > 0$ on the branch $\mathcal{I}m\ (\lambda - \frac{1}{4})^{1/2} > 0$ is necessary to insure a unique $G(r, r')$ for this non-Hermitean case.

The homogeneous solutions $t(r, r_2)$ and $u(r, r_2)$ satisfying the continuity conditions at r_2 and boundary conditions at $r = 0$ and ∞ are again given by equations (4.39) and (4.40). The reactances X_ν^- and X_ν^+ at r_2 are the same as in equations (4.44) and (4.47), respectively. The characteristic Green's function is of the form shown in equation (4.41). As noted above X_ν^- has a branch point at $\lambda = \frac{1}{4}$ whereas X_ν^+ is regular. Hence $G(r, r')$ possesses singularities in the λ plane in the form of a branch cut along the λ axis from $\lambda = \frac{1}{4}$ to ∞ and simple poles at the complex zeros ν_i, if any, of X_ν. The spectrum for this case thus may be both discrete and continuous.

Evaluation of the completeness relation (3.8) requires the integration of the characteristic Green's function along a contour enclosing both the poles and the branch cut of the latter. The result may be given in the form shown in equation (4.45) from which on insertion of the relevant expressions for X_ν^- and X_ν^+ the discrete eigenfunctions are seen to be

$$(4.49) \qquad \begin{aligned} &\left[\frac{2\nu_i + 1}{(\partial X_\nu/\partial \nu)_{\nu_i}}\right]^{1/2} \frac{j_{\nu_i}(k_1 r)}{j_{\nu_i}(k_1 r_2)}, \qquad 0 < r < r_2 \\ &\left[\frac{2\nu_i + 1}{(\partial X_\nu/\partial \nu)_{\nu_i}}\right]^{1/2} \frac{h_\nu^{(1)}(k_2 r)}{h_\nu^{(1)}(k_2 r_2)}, \qquad r_2 < r \end{aligned}$$

with the ν_i defined as the roots of

$$X_\nu = \frac{k_1}{\epsilon_1} \frac{j_\nu'(k_1 r_2)}{j_\nu(k_1 r_2)} + \frac{k_2}{\epsilon_2} \frac{h_\nu^{(1)\prime}(k_2 r_2)}{h_\nu^{(1)}(k_2 r_2)} = 0.$$

The discontinuously represented and biorthogonal eigenfunctions of the continuous spectrum likewise follow from (4.45) as

$$(4.50) \qquad \begin{aligned} &\left[\frac{2\nu + 1}{X_\nu}\right]^{1/2} \frac{j_\nu(k_1 r)}{j_\nu(k_1 r_2)}, \qquad 0 < r < r_2 \\ &\left[\frac{2\nu + 1}{X_\nu}\right]^{1/2} \frac{h_\nu^{(1)}(k_2 r)}{h_\nu^{(1)}(k_2 r_2)}, \qquad r_2 < r, \end{aligned}$$

where $+0 - i\infty < \nu + \frac{1}{2} i < +0 + i\infty$ defines the continuous spectrum (note that a $1/(2\pi i)$ factor is understood in superposing eigenfunctions). An alternative symmetric orthogonal form for the eigenfunctions of the continuous spectrum is also obtainable as in equation (3.21b); in this case the eigenfunctions are proportional to $u(r, r_2)$.

5. *Field of a Point Source. Multi-dimensional Green's Functions*

As noted in Section 2 the determination of the electromagnetic field produced by an arbitrary vector current distribution may be reduced to two scalar problems for the potentials. The latter are defined by inhomogeneous partial differential equations which in turn may be reduced to ordinary one dimensional differential equations on introduction of an appropriate set of eigenfunctions, or modes. Since the field of a known current distribution follows by integration from the corresponding field of a point or dipole source it is sufficient to treat only the latter problem (cf. Section 3b), i.e., if we exclude diffraction problems in which the current distribution is not known and also has to be determined. For definiteness the discussion below will refer to the field of a vertical (radial) electrical dipole, which excites only E type modes. The H type field of a vertical magnetic dipole follows by duality; however, the field of an arbitrarily oriented electric or magnetic dipole entails a somewhat greater degree of complexity because of the excitation of both E and H type modes, but nevertheless follows by similar methods (cf. Section 2). The nature of the spherically stratified region in which, for illustration, it is desired to obtain a representation of the field is shown in Figure 5.1. The examples to be discussed are distinguished

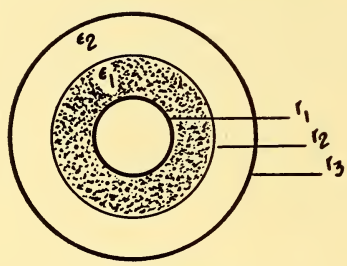


FIGURE 5.1

by the values ascribed to the radii r_1, r_2, r_3 of the boundaries (if any) and to the relative dielectric constants ϵ_1 and ϵ_2 (for simplicity, $\mu_1 = \mu_2 = 1$) of the strata.

The various representations of a dipole field are clearly exhibited by phrasing the relevant mathematical problem as a Green's function problem. The choice of a vertical dipole implies that only two dimensional problems will be considered. The Green's function of interest then satisfies the inhomogeneous equation[17]

$$\left(\frac{\partial}{\partial r}\frac{1}{\epsilon}\frac{\partial}{\partial r} + \frac{1}{\epsilon r^2}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + k^2\mu\right)\mathcal{G}(r, r') = -\delta(r - r')\frac{\delta(\theta - \theta')}{\sin\theta'} \tag{5.1}$$

with boundary conditions and variability of ϵ and μ to be stated. Equation (5.1)

[17]Strictly stated, the azimuthal symmetry of the field of a single vertical (radial) dipole obtains only for $\theta' = 0$. To emphasize the symmetry of the mathematical problem θ' will be retained explicitly; this corresponds to a ring of radial electric dipoles.

is evidently a special case of equation (2.25b). As discussed in Section 2, equation (5.1) represents but one of a variety of ways of phrasing the field problem for a vertical electric dipole. One virtue of the Green's function formulation is that there exists a formal prescription for solving equation (5.1) in terms of the eigenfunctions or characteristic Green's function of the component one dimensional operators in (5.1). Thus, as discussed in Sections 3 and 4, let the eigenfunctions, eigenvalues, and characteristic Green's functions of the two relevant operators in (5.1) be defined by

$$\left(\frac{d}{d\theta}\sin\theta\frac{d}{d\theta} + \hat{\lambda}_i \sin\theta\right)\hat{\phi}_i(\theta) = 0 \tag{5.2a}$$

$$\left(\frac{d}{d\theta}\sin\theta\frac{d}{d\theta} + \hat{\lambda}\sin\theta\right)\hat{G}(\theta, \theta'; \hat{\lambda}) = -\delta(\theta - \theta') \tag{5.2b}$$

and

$$\left(\frac{d}{dr}\frac{1}{\epsilon}\frac{d}{dr} + k^2\mu + \frac{\lambda_i}{\epsilon r^2}\right)\phi_i(r) = 0 \tag{5.2c}$$

$$\left(\frac{d}{dr}\frac{1}{\epsilon}\frac{d}{dr} + k^2\mu + \frac{\lambda}{\epsilon r^2}\right)G(r, r'; \lambda) = -\delta(r - r'). \tag{5.2d}$$

The boundary conditions on $\hat{G}$ and G correspond to those on $\mathcal{G}(r, r')$ at the endpoints of the θ and r intervals, respectively; the conditions on $\hat{\phi}_i$ and ϕ_i are given in terms of those on $\hat{G}$ and G by relations of the form (3.3b). In terms of the above defined functions the completeness relation (3.8) is given in the θ interval by

$$\frac{\delta(\theta - \theta')}{\sin\theta'} = -\frac{1}{2\pi i}\int_{\hat{C}} \hat{G}(\theta, \theta'; \hat{\lambda})\, d\hat{\lambda} = \sum \hat{\phi}_i(\theta)\hat{\phi}_i(\theta'), \tag{5.3a}$$

and in the r interval by

$$\epsilon r'^2 \delta(r - r') = -\frac{1}{2\pi i}\int_{\hat{C}} G(r, r'; \lambda)\, d\lambda = \sum_i \phi_i(r)\phi_i(r'), \tag{5.3b}$$

where the contour $\hat{C}$ is extended about all the singularities of $\hat{G}$ in the complex $\hat{\lambda}$ plane, and where C encloses all the singularities of G in the λ plane. With the knowledge of equations (5.2-3) the solution to equation (5.1) may be represented immediately in the alternative forms

$$\mathcal{G}(r, r') = \sum_i G(r, r'; -\hat{\lambda}_i)\hat{\phi}_i(\theta)\hat{\phi}_i(\theta') \tag{5.4a}$$

$$= -\frac{1}{2\pi i}\int_{\hat{C}} G(r, r'; -\hat{\lambda})\hat{G}(\theta, \theta'; \hat{\lambda})\, d\hat{\lambda} \tag{5.4b}$$

$$= -\frac{1}{2\pi i}\int_C \hat{G}(\theta,\,\theta';\,-\lambda)G(r,\,r';\,\lambda)\;d\lambda \tag{5.4c}$$

$$= \sum_i G(\theta,\,\theta';\,-\lambda_i)\phi_i(r)\phi_i(r')\ ^{18} \tag{5.4d}$$

of which the first two are so called representations in θ, the latter two in r. Equations (5.4) may be verified by noting that the operation

$$\frac{\partial}{\partial r}\frac{1}{\epsilon}\frac{\partial}{\partial r} + k^2\mu + \frac{1}{\epsilon r^2}\frac{1}{\sin\,\theta}\frac{\partial}{\partial\theta}\sin\,\theta\,\frac{\partial}{\partial\theta}$$

under the sum (integral) sign of each of the right hand members in (5.4) yields the desired result: $-\delta(r - r')\delta(\theta - \theta')/\sin\,\theta'$. The interchange of summation (integration) and differentiation is permitted in view of the identity of the boundary conditions on $\mathcal{G}\,(r,\,r')$ and on $G(r,\,r')$ and $\hat{G}(\theta,\,\theta')$.

Equations (5.4a-d) provide four different representations of the Green's function defined in (5.1), each representation having different convergence properties depending on the range involved. For example: in far field calculations of plane wave scattering from small (compared to a wavelength) spherically stratified structures equation (5.4a) is most convergent; in far field calculations when the wavelength is small—the geometrical optical limit—equations (5.4b) or (5.4c) are most useful; in the calculation of the far surface field of a dipole located near the surface of a large spherical structure the representation (5.4d) is most convergent. From any one of the representations in (5.4) the others of (5.4) may be derived by appropriate sum—integral equivalences or contour deformations (cf. Section 3). Thus, Watson's classical treatment of the spherical earth and allied problems entails a transition from the representation in (5.4a) to that in (5.4d) via the intervening representations. The desirability of selecting the relevant representation without intervening function theoretic considerations should be evident.

The characteristic one dimensional representations obtained in Section 4 will now be employed to obtain alternative representations of the vertical electric dipole field in the stratified region $r_1 < r < r_3$, $0 < \theta < \pi$, shown in Figure 5.1. Thus, we seek a solution of equation (5.1) subject to conditions of finiteness at $\theta = 0,\,\pi$ and to various, as yet unstated, boundary conditions at $r = r_1$, r_3. In view of the representations (4.4) and (4.6) in θ, and of the representations (4.41) and (4.42) in r, the solutions (5.4) may be written more explicitly in the notation of Sec. 4c as

$$\mathcal{G}(r,\,r') = \sum_{n=0}^{\infty}\frac{t_n(k_\alpha r_<,\,k_\alpha r_2)\,u_n(k_\alpha r_>,\,k_\alpha r_2)}{-X_n}\left(n + \frac{1}{2}\right)P_n(\cos\,\theta)P_n(\cos\,\theta') \tag{5.5a}$$

$$= \frac{1}{2\pi i}\int \frac{t_\nu(k_\alpha r_<,\,k_\alpha r_2)\,u_\nu(k_\alpha r_>,\,k_\alpha r_2)}{-X_\nu}\cdot\frac{P_\nu(\cos\,\theta_<)P_\nu(-\cos\,\theta_>)}{(2/\pi)\,\sin\,\nu\pi}\,d\hat{\lambda} \tag{5.5b}$$

[18]See footnote 8.

$$= \frac{1}{2\pi i} \int_C \frac{P_\nu(\cos\theta_<)P_\nu(-\cos\theta_>)}{-(2/\pi)\sin\nu\pi} \cdot \frac{t_\nu(k_\alpha r_>, k_\alpha r_2)\, u_\nu(k_\alpha r_>, k_\alpha r_2)}{-X_\nu}\, d\lambda \tag{5.5c}$$

$$= \sum_i \frac{P_{\nu_i}(\cos\theta_<)P_{\nu_i}(-\cos\theta_>)}{-(2/\pi)\sin\nu_i\pi} \frac{2\nu_i + 1}{(\partial X_\nu/\partial\nu)_{\nu_i}}$$

$$\cdot t_{\nu_i}(k_\alpha r, k_\alpha r_2)\, t_{\nu_i}(k_\alpha r', k_\alpha r_2) \tag{5.5d}$$

$$+ \frac{1}{2\pi i} \int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} \frac{P_\nu(\cos\theta_<)\, P_\nu(-\cos\theta_>)}{(2/\pi)\sin\nu\pi} \frac{t_\nu(k_\alpha r, k_\alpha r_2)\, u_\nu(k_\alpha r', k_\alpha r_2)}{X_\nu}$$

$$\cdot (2\nu + 1)\, d\nu$$

where

$$t_\nu(x, y) = c_\nu(x, y) - X_\nu^- \frac{s_\nu(x, y)}{k_\alpha/\epsilon_\alpha}$$

$$u_\nu(x, y) = c_\nu(x, y) + X_\nu^+ \frac{s_\nu(x, y)}{k_\alpha/\epsilon_\alpha}$$

and where the index α is 1 or 2 depending, respectively, on whether r (or r') is in the range r_1 to r_2 in which ϵ_α is ϵ_1 or the range r_2 to r_3 in which ϵ_α equals ϵ_2. It should be noted that since $\lambda = -\nu(\nu + 1) = -\hat{\lambda}$, the contours C and $\hat{C}$ are simply related when drawn in the same λ, or $\hat{\lambda}$, or ν plane. Hence the representations (5.5b) and (5.5c) differ only slightly from one another; the transition between these two representations involves, however, certain function theoretic properties of the integrand which need not be adduced in the direct application of the representation theorems (5.5b) or (5.5c). It is evident that the form (5.5d) corresponds in general to the existence of both a discrete and continuous set of eigenfunctions in r.

In the following we shall confine ourselves to a more detailed discussion of the representation (5.5d) which expresses the vertical dipole field in terms of waves guided along the θ-direction. It is this form which is most convergent in many spherical propagation problems. The characteristic representations developed in Sections 4b and c provide explicit values for X_ν^- and X_ν^+ in a number of stratified regions. Since it is of more interest to indicate the detailed nature of the spectrum rather than the detailed structure of the field in these cases, the latter will be omitted in the following. Detailed discussions of field structure for $\epsilon(r)$ variations encountered in practice are contained in the works of Bremmer (loc. cit., Chaps. VIII to X) and Booker and Walkinshaw (loc. cit.). In parallel with the discussion in Sections 4b and c the following cases are considered in more detail:

(1) $\partial\mathcal{G}/\partial r = 0$ *at* $r = r_1$ *and* r_3

This case is of interest in connection with an idealized tropospheric layer of constant dielectric constant ϵ_1 extending from r_1 to r_2 and an isotropic ionospheric

layer of constant dielectric constant ϵ_2 from r_2 to r_3 ; these layers are bounded below by a perfectly conducting earth at r_1 and above by a perfectly conducting sheath at r_3 . The nature of the propagation is dependent on the relative values of ϵ_1 and ϵ_2 .

Since X_ν is a meromorphic function in this case, there is only a discrete spectrum. On insertion of the values for X_ν^- and X_ν^+ from equations (4.39a) and (4.40a), equation (5.5d) becomes for $r_1 < \overset{r}{\underset{r'}{}} < r_2$ and $0 < \theta < \pi(\theta' = 0)$

$$\mathcal{G}(r, r') = \sum_i \frac{P_{\nu_i}(-\cos\theta)}{-(2/\pi)\sin\nu_i\pi} \frac{(2\nu_i + 1)}{(\partial X_\nu/\partial\nu)_{\nu_i}} \frac{c_{\nu_i}(k_1 r, k_1 r_1)}{c_{\nu_i}(k_1 r_2\,, k_1 r_1)} \frac{c_{\nu_i}(k_1 r', k_1 r_1)}{c_{\nu_i}(k_1 r_2\,, k_1 r_1)} \tag{5.6}$$

with ν_i determined by $X_{\nu_i} = X_{\nu_i}^- + X_{\nu_i}^+ = 0$. For large kr the latter equation can be approximated by a saddle point or by a W.K.B. procedure as

$$\frac{k_1}{\epsilon_1}\sin\beta_1 \tan\left[\gamma_1(r_2) - \gamma_1(r_1)\right] + \frac{k_2}{\epsilon_2}\sin\beta_2\tan\left[\gamma_2(r_3) - \gamma_2(r_2)\right] = 0 \tag{5.7}$$

where

$$\cos\beta_\alpha = \frac{\nu_i + \frac{1}{2}}{k_\alpha r_2} = \frac{\zeta_i}{k_\alpha} \tag{5.8}$$

$$\gamma_\alpha(r) = k_\alpha r\,(\sin\beta_\alpha - \beta_\alpha\cos\beta_\alpha) + \frac{\pi}{4}$$

with $\alpha = 1$ or 2. If $h_\alpha/r_2 \ll 1$, where $h_1 = r_2 - r_1$ and $h_2 = r_3 - r_2$, equation (5.7) can be replaced by the "parallel plane" approximation

$$\begin{aligned}&[(k^2\epsilon_1 - \zeta_i^2)^{1/2}/\epsilon_1]\tan(k^2\epsilon_1 - \zeta_i^2)^{1/2}h_1 \\ &\qquad + [(k^2\epsilon_2 - \zeta_i^2)^{1/2}/\epsilon_2]\tan(k^2\epsilon_2 - \zeta_i^2)^{1/2}h_2 = 0\end{aligned} \tag{5.9}$$

provided $k_\alpha \equiv k(\epsilon_\alpha)^{1/2} \neq \zeta_i$. The zeros of (5.9) can be readily found and, when ϵ_α is real, they yield for $\lambda_i = -\nu_i(\nu_i + 1)$ real values of two distinct types: those for which $\lambda_i - \frac{1}{4} > 0$, i.e. $\nu_i + \frac{1}{2}$ imaginary, and those for which $\lambda_i - \frac{1}{4} < 0$, i.e. $\nu_i + \frac{1}{2}$ real. The former values characterize modes that attenuate in the θ-direction,[19] whereas the latter correspond to modes propagating in the θ-direction. There are only a finite number of propagating modes (for finite $r_3 - r_1$) and these can be further distinguished by whether ζ_i lies between $k(\epsilon_1)^{1/2}$ and $k(\epsilon_2)^{1/2}$ or by whether ζ_i is less than both $k(\epsilon_1)^{1/2}$ and $k(\epsilon_2)^{1/2}$. The modes for which $k(\epsilon_2)^{1/2} < \zeta_i < k(\epsilon_1)^{1/2}$ are so called "surface waves" or "trapped waves." They correspond to waves having no r propagation (attenuated) in

[19]Note that for ν_i large and with a positive imaginary part, and for $1/\nu_i < \theta < \pi - 1/\nu_i$,

$$\frac{P_{\nu_i}(-\cos\theta)}{-(2/\pi)\sin\nu_i\pi} \approx \left(\frac{\pi i}{2}\right)^{1/2} \frac{\exp\{i(\nu_i + \frac{1}{2})\theta\}}{[(\nu_i + 1)\sin\theta]^{1/2}}\,.$$

the dielectric ϵ_2 and suffering internal reflection in dielectric ϵ_1 — or conversely if $\epsilon_2 > \epsilon_1$.

The non-Hermitean case with complex ϵ_α yields complex values for λ_i, but save for the additional attenuation introduced thereby this case is for the most part the same as for real ϵ_α. In the limit case $\epsilon_1 = \epsilon_2$ there are no surface waves

(2) $\partial \mathcal{G}/\partial r = 0$ *at* r_3, $\mathcal{G}$ *finite at* $r_1 = 0$.

This case sheds an interesting light on the effect of a spherical earth of dielectric constant ϵ_1 and radius r_2 upon propagation in the presence of a perfectly conducting ionospheric sheath at r_3.

On insertion into equation (5.5d) of the relevant values for X_ν^- and X_ν^+ from equations (4.44) and (4.40a), one obtains in virtue of the non-analytic character of X_ν^- both a discrete and continuous spectrum for this case. The continuous modes correspond to the coalescence of the discrete modes with $\lambda_i - \frac{1}{4} > 0$ of the preceding example. On the other hand the discrete modes for this case are found for those values $\lambda_i = -\nu_i(\nu_i + 1) < \frac{1}{4}$ that satisfy the resonance equation $X_{\nu_i} = X_{\nu_i}^- + X_{\nu_i}^+ = 0$ or, with the same approximation and notation as in equations (5.8-9),

$$-\frac{k_1}{\epsilon_1}\sin\beta_1\tan\gamma_1(r_2) + [(k^2\epsilon_2 - \zeta_i^2)^{1/2}/\epsilon_2]\tan(k^2\epsilon_2 - \zeta_i^2)^{1/2}h_2 = 0. \tag{5.10}$$

Again only a finite number of such modes are admitted if kr_3 is finite although of course this number may be large. Surface waves are possible both for $\epsilon_1 > \epsilon_2$ and for $\epsilon_2 > \epsilon_1$, in the latter case an inverted surface wave appears.

The non-Hermitean case with complex ϵ_α, as in the previous example, requires a slight modification of the above results in that all the ν_i are complex

(3) $\partial \mathcal{G}/\partial r = 0$ *at* r, $\mathcal{G} \to 0$ *as* $r_3 \to \infty$

The idealization involved in this case appears suitable for the description of anomalous propagation in the presence of a tropospheric layer of dielectric constant ϵ_1 and height $h_1 = r_2 - r_1$ bounded below by a perfectly conducting spherical earth and above by an atmosphere of dielectric constant ϵ_2. It will be assumed that $\epsilon_2/\epsilon_1 < 1$.

The meromorphic reactances X_ν^- and X_ν^+ are given in this instance by equations (4.39a) and (4.47). The associated spectrum is discrete and in the range $r_1 < \frac{r}{r'} < r_2$ (with $\theta' = 0$) the representation (5.5d) becomes the same as in equation (5.6) where the ν_i are again the roots of $X_{\nu_i} = X_{\nu_i}^- + X_{\nu_i}^+ = 0$. These roots are either complex or real (if ϵ_α is real). With the same approximations and notation as in equations (5.8-9) the complex ν_i are determined by

$$[(k^2\epsilon_1 - \zeta_i^2)^{1/2}/\epsilon_1]\tan(k^2\epsilon_1 - \zeta_i^2)^{1/2}h_1 + (k_2/\epsilon_2)\sin\beta_2\tan\gamma_2(r_2) = 0. \tag{5.11}$$

For numerical convenience one defines

$$\nu_i = k_2 r_2 + (k_2 r_2)^{1/3}\tau_i \tag{5.12}$$

To the indicated approximation $\gamma_2(r_2) \approx k_2 r_2\, \beta^3/3 \approx [(-2\tau_1)^{3/2}]/3$ and hence equation (5.12) becomes

$$[(k_1^2 - k_2^2)^{1/2}/\epsilon_1] \tan (k_1^2 - k_2^2)^{1/2} h_1 + \frac{k_2}{\epsilon_2}\frac{(-2\tau_i)^{1/2}}{(k_2 r_2)^{1/3}} \tan \left[\frac{\pi}{4} + \frac{1}{3}(-2\tau_i)^{3/2}\right] = 0, \tag{5.14}$$

from which an infinite set of τ_i, and hence ν_i, can be found. The numbers τ_i so found are closely related to those encountered in the case of an imperfectly conducting earth (cf. equation (5.16) below). In the range where real roots ν_i are possible $\tan \gamma_2(r_2) \approx i$; hence the equation $X_{\nu_i} = 0$ has by (5.11) the "parallel plane" approximation

$$[(k^2\epsilon_1 - \zeta_i^2)^{1/2}/\epsilon_1] \tan (k^2\epsilon_1 - \zeta_i^2)^{1/2} h_1 - [(\zeta_i^2 - k^2\epsilon_2)^{1/2}/\epsilon_2] = 0. \tag{5.15}$$

Equation (5.15) admits only real roots ζ_i which are finite in number if kh_1 is finite. They characterize "surface waves" or "trapped modes" propagating in the θ direction. The surface waves encountered in example 1 differ only slightly from those for this case, as is evident from the fact that in the relevant range of ζ_i the right hand tangent in equation (5.10) is approximately unity. This is indicative of the near independence of the surface waves on the nature of the boundary surface at $r_3 \to \infty$. A feature in marked contrast with that of the waves, defined by (5.11), which are radically affected by the nature of the boundary at r_3. The latter fact is intimately related to the "limit circle" character of the singular point r_3 in Hermitean case.

As in the previous example the case of complex ϵ_α introduces minor modifications. The limiting case $\epsilon_1 = \epsilon_2$ likewise admits complex ν_i, as is seen from equation (5.14), but there are no surface waves. For $\epsilon_1 = \epsilon_2$ the functions $c_{\nu_i}(k_1 r, k_1 r_1)/c_{\nu_i}(k_1 r_2, k_1 r_1)$ in (5.11) should be replaced by $h_{\nu_i}^{(1)}(k_1 r)/h_{\nu_i}^{(1)}(k_1 r_1)$.

(4) $\mathcal{G}$ *finite at* $r_1 = 0$, $\mathcal{G} \to 0$ $r_3 \to \infty$

This is the oft treated problem of a spherical earth of dielectric constant ϵ_1 and radius r_2 bounded above by an atmosphere of dielectric constant $\epsilon_2(\epsilon_2 < \epsilon_1)$.

The reactances X_ν^- and X_ν^+ are given in equations (4.44) and (4.47). The non-analyticity of X_ν^- implies that the spectrum for this case is both discrete and continuous. The ν_i for the discrete modes are determined as always by the zeros of the equation $X_{\nu_i} = X_{\nu_i}^- + X_{\nu_i}^+ = 0$. The latter can be approximated for large $k_2 r_2$ by replacing $X_{\nu_i}^-$ by the left hand member of equation (5.10) and $X_{\nu_i}^+$ by the right hand member of equation (5.11). On introduction of the parameters τ_i defined in (5.13) and use of the approximation $\tan \gamma_1(r_2) \approx i$ (i.e. assuming a slight amount of dissipation) this equation becomes

$$(5.16)\qquad -\frac{i(k_1^2 - k_2^2)^{1/2}}{\epsilon_1} + \frac{k_2}{\epsilon_2}\frac{(-2\tau_i)^{1/2}}{(k_2 r_2)^{1/3}} \tan\left[\frac{\pi}{4} + \frac{1}{3}(-2\tau_i)^{3/2}\right] = 0.$$

Equation (5.16), which is treated in detail by Bremmer (loc. cit. Eq. (21b) p. 43), admits an infinite number of complex zeros τ_i. The associated ν_i correspond to the modes defined by equation (5.12) of the previous problem. It is of interest to note that there is no discrete surface wave in this case, as is also true in the plane earth limit $r_2 \to \infty$.

(5) *Conical Region.* $\partial \mathcal{G}/\partial\theta = 0$ *at* $\theta = \theta_1$

The mode representations just discussed are applicable to a variety of other cases. For example, if the spherically stratified regions of the preceding examples are bounded by a conducting wall at $\theta = \theta_1$, one has only to replace in (5.5d) the Green's function (4.4) by (4.9) to obtain the field of a radially directed, h-independent electric dipole distribution. The consequent change in the nature of the θ-propagation represents the sole modification in the results discussed above.

Bibliography

1. Bremmer, H., *Terrestrial Radio Waves*, Elsevier Publishing Company, New York–Houston, 1949, Chapters II–VI.
2. Weyl, H., *Über gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen*, Mathematische Annalen, Volume 68, 1910, p. 220.
3. Titchmarsh, E. C., *Eigenfunction Expansions*, Oxford Univ. Press, 1946, Chapter II.
4. Sommerfeld, A., *Electromagnetische Schwingungen*, in Frank, P., and von Mises, R., *Die Differential und Integralgleichungen der Mechanik und Physik*, Volume II, Rosenberg, New York, 1943, p. 891.
5. Booker, H. G., and Walkinshaw, W., *The mode theory of tropospheric refraction and its relation to wave-guides and diffration.* Meteorological Factors in Radio Wave Propagation. Report of a conference held on April 8, 1946, by the Physical Society and The Royal Meteorological Society. The Physical Society, London, 1946, p. 80.
6. Friedrichs, K. O., *Lectures on spectral representations of linear operators*, New York University, New York, 1948.
7. Stratton, J. A., *Electromagnetic Theory*, McGraw-Hill, New York, 1941.
8. Lebedev, N. N., *On the representation of a function as an integral of a MacDonald function*, Dokl. Acad. Nauk. U. S. S. R., Volume 58, No. 6, 1947, p. 1007.

Propagation in a Non-homogeneous Atmosphere*

By BERNARD FRIEDMAN
Washington Square College of Arts and Science, New York University

I. *Introduction*

The problem of explaining the propagation of radio waves around the surface of the earth has long been of great interest to both physicists and mathematicians because of its practical importance and because of its considerable mathematical difficulties.

There are two fundamental approaches to the theory of such propagation. The first approach, which is that of Watson [1], van der Pol and Bremmer [2], and Rydbeck [3], and which has been applied mostly to ionospheric problems, assumes the earth to be a perfect sphere with finite or infinite conductivity, and expresses the field due to a radiating dipole as an infinite series involving spherical Bessel functions. Since the series converges much too slowly for practical purposes, it is transformed into a contour integral which can be evaluated by the method of residues. To find the residues of this integral, highly complicated asymptotic expansions for the spherical Bessel functions must be used. These asymptotic expansions, which are derived by the method of stationary phase, involve the Hankel functions of order 1/3 or the Airy integral, both of which occur frequently in diffraction theory.

The second approach to the problem is that used by Pryce [4] and Booker [5] in England, and by Furry [6] and Pekeris [7] in the United States while studying the effect of the troposphere on radio wave propagation. In this approach, the earth is assumed to be flat but, following a suggestion of Schelling and Burrows [8], the index of refraction is modified so as to take into account the curvature of the earth. The field due to a radiating dipole is now expressed as an infinite integral involving ordinary Bessel functions. This integral, also, can be evaluated by the method of residues but it was realized that the residues are really the successive terms in an eigenfunction expansion for the field of the

---

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

---

*This work was performed at Washington Square College of Arts and Science, New York University and was supported in part by Contract No. AF-19(122)-42 with the U.S. Air Force through sponsorship of the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

dipole. The propagation problem is therefore reduced to the study of an eigenvalue problem for an ordinary differential equation. In case the atmosphere above the earth is assumed to have a constant refractive index, the eigenfunction expansion is in terms of Hankel functions of order 1/3, exactly the same functions which appear in the first approach.

In this paper we start with Maxwell's equations and then present a complete exposition of the theory of propagation in an inhomogeneous atmosphere where the dielectric constant is stratified in a radial direction. The theory can be applied to both ionospheric and tropospheric problems. The methods we have used and the results we have obtained are similar to those contained in the papers by Watson [1], van der Pol and Bremmer [2], and in the books by Bremmer [11] and Rydbeck [3]. The novelties of our treatment and results are the following:

(1) The dielectric constant is considered to be a single function of position instead of different functions in different regions of space. As a result, the formulas for the Hertz potential and the electro-magnetic field have a simpler appearance and are not as complicated and clumsy as in the usual treatment. All the necessary transformations and approximations can be carried out on these formulas and because of their simpler appearance it is much easier to acquire an insight into their meaning. Of course, when quantitative results are wanted, it is necessary to consider the dielectric constant as composed of different functions in different regions, to solve the appropriate differential equation in each region and then to match the solutions across the boundary to obtain the complete solution.

(2) The expression of the Hertz potential as a contour integral and the transformation of the contour integral into a sum of residues are carefully discussed. There are two questions in this procedure which are usually ignored. One is the question of whether the contour integral over an infinite semi-circle may be neglected. Watson [1] has shown that in the case of a constant atmosphere these integrals vanish. However, his proof depends upon the use of asymptotic formulas for the Bessel functions. We show in the Appendix that, even for a completely arbitrary atmosphere, these integrals vanish. The second question is this: In general, the contour integral can be expressed as a sum of residues plus an integral around the imaginary axis; when can this integral be neglected? Watson [1] shows that, in the case of the constant atmosphere, this integral is negligible compared to the residue series. Later writers on this topic either drop the integral without comment, or change the boundary condition at the center of the earth [11], or express the Hertz potential in terms of incident and reflected waves [2], which result only in a residue series and not in an integral. This latter procedure can be applied to the general case, but it is only in the case of the constant atmosphere that the integral over the imaginary axis does not appear. In the appendix we consider an arbitrary atmosphere and show, first, that if the earth is assumed to be a perfect conductor, the integral over the imaginary axis vanishes identically; and, second, that if the earth

is not assumed to be a perfect conductor but if its dielectric constant is assumed to be large compared to that of the atmosphere, which is usually the case in practice, then the integral over the imaginary axis is small compared to the first term of the residue series.

(3) The general nature of our study emphasizes the fact that the propagation problem reduces to the solution of the eigenvalue problem for an ordinary differential equation. This fact, while it is clearly recognized in the flat earth theory and while it is implicitly contained in most previous work, has heretofore not been explicitly stated for a spherical earth. Once this fact is known, it is clear that the W.K.B. method should be applied. By means of this method, the "phase integral" method of Eckersley [10] can be understood and made more exact.

(4) It is shown that the flat earth theory follows from the exact spherical earth theory by making suitable approximations in the ordinary differential equation mentioned in (3). From the nature of the approximations, it is concluded that the eigenvalues will be closely approximated but that the expressions for the field on the flat earth theory will be inaccurate at large distances above the earth. These conclusions agree with those of Pekeris [7] in his analysis of the accuracy of the flat earth approximation.

(5) In the case of tropospheric propagation the W.K.B. method gives explicit results. First, if the atmosphere is inhomogeneous but does not have a duct, the field can be obtained by a change of scale from that produced by a constant atmosphere. This result generalizes a similar result obtained by Eckersley [10] and Bremmer [11] for a special form of variation of the dielectric constant. Second, if the atmosphere does have a duct, the differential equation that must be considered will have two transition points and its solution will depend on confluent hypergeometric functions. We find that the differential equation can be approximated by one which has a double transition point and that then the solution can be expressed in terms of Hankel functions of order one-fourth. This approach seems to be completely new and suggests a simple, direct quantitative method for approach to the problem of propagation in a duct.

The problem of propagation in an inhomogeneous atmosphere is formulated mathematically in section 2 and it is shown that the problem requires the solution of a scalar partial differential equation. In Section 3 this equation is solved by the classical method of separation of variables and expansion in terms of Legendre polynomials. In Section 4 this solution is represented as a contour integral and evaluated by the calculus of residues. These residues are interpreted in Section 5 as an expansion of the solution in terms of radial eigenfunctions. In Section 6 we use the fact that the wavelength of the radiation is very small compared to the radius of the earth to obtain simpler boundary conditions at the earth's surface and a simpler expression for the solution. We return in Section 7 to the discussion of the radial eigenfunction and we show how, by the use of Langer's extension of the W.K.B. method, useful approximations to the eigenfunctions can be found.

The case of a uniform atmosphere is discussed by this method in Section 8 and the well known results are obtained. Section 9 shows that the case of the non-uniform atmosphere without ducts can be treated almost as easily as the case of a uniform atmosphere. In Section 10 the "flat earth" theory and the modified index of refraction are shown to be reasonable approximations to the exact theory. Section 11 discusses the case of an atmosphere with ducts; Section 12 summarizes the work for the electric dipole. The detailed mathematical questions concerning the contour integral representation are discussed in the appendices.

2. *Formulation of the Problem*

Assume the earth is a sphere, radius a, with complex dielectric constant ϵ_1. Introduce spherical polar coordinates, r, θ, ϕ, where r is the distance from the center of the earth, θ is the co-latitude and ϕ the longitude. We assume that the earth is immersed in a medium whose dielectric constant, ϵ, is a function of r alone. The propagation problem depends on finding the field produced by a

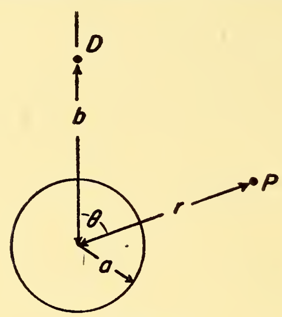


FIGURE 1

unit dipole which is radiating at a constant angular frequency ω and is located at a distance, b, from the center of the earth. By a suitable rotation of the coordinate system we can always locate the dipole on the polar axis as indicated in the above diagram.

The electromagnetic field produced by the dipole should satisfy Maxwell's equations, both inside and outside the earth, and should possess at the point D a singularity corresponding to that of a dipole while on the earth's surface the tangential components of the electric field **E** and the magnetic field **H** must be continuous.

Let **E** exp $\{-i\omega t\}$ represent the vector electric field intensity and **H** exp $\{-i\omega t\}$ the vector magnetic field intensity produced by the dipole, then Maxwell's equations in M.K.S. units become the following:

$$\begin{aligned} &\nabla \times \mathbf{E} = i\omega\mu\mathbf{H}, \qquad \nabla \cdot \mathbf{H} = 0 \\ &\nabla \times \mathbf{H} = -i\omega\epsilon\mathbf{E}, \qquad \nabla \cdot \epsilon\mathbf{E} = 0. \end{aligned} \tag{1}$$

Here, $\epsilon = \epsilon(r)$, is a scalar function which represents the variation of the dielectric constant of the medium. For $r < a$, $\epsilon(r)$ is assumed to be a constant, ϵ_1 representing the complex dielectric constant of the earth.

We shall derive the well known fact [12] that in spherical coordinates the electromagnetic field given by (1) can be obtained from a Hertz vector which has only its radial component non-zero, even when the dielectric constant is a function of r. Since $\epsilon\mathbf{E}$ is a vector whose divergence is zero, we may write

$$\epsilon\mathbf{E} = i\omega\mu\,\nabla \times \epsilon\mathbf{C}. \tag{2}$$

Substituting this expression for $\mathbf{E}$ in the curl $\mathbf{H}$ equation, we get

$$\nabla \times (\mathbf{H} - \omega^2\mu\epsilon\mathbf{C}) = 0$$

so that the expression in parenthesis must be the gradient of some scalar, which we call ψ. We have then

$$\mathbf{H} = \omega^2\mu\epsilon\mathbf{C} + \nabla\psi. \tag{3}$$

Using (2) and (3) in the remaining equations of (1), we obtain the following equations for $\mathbf{C}$ and ψ:

$$\nabla \times \frac{1}{\epsilon}(\nabla \times \epsilon\mathbf{C}) - \omega^2\mu\epsilon\mathbf{C} - \nabla\psi = 0 \tag{4}$$

$$\nabla \cdot (\omega^2\mu\epsilon\mathbf{C}) + \nabla^2\psi = 0. \tag{5}$$

Note that if we take the divergence of equation (4), we obtain equation (5) so that it is sufficient to satisfy equation (4).

Let us express these equations in spherical coordinates. Since ϵ is a function of r only and since the dipole is on the polar axis, the electromagnetic field and therefore also the Hertz vector $\mathbf{C}$ will not depend on the longitude ϕ but only on the coordinates r, θ. We assume that the vector $\mathbf{C}$ has a radial component of the form $rU(r, \theta)$ where $U(r, \theta)$ is some scalar function, and that its other components are zero. The θ-component of (4) implies that

$$\frac{\partial\psi}{\partial\theta} = \frac{\partial^2}{\partial r\,\partial\theta}(rU)$$

so that we may take $\psi = \partial(rU)/\partial r$ and then the r-component of (4) reduces exactly to the following:

$$\Delta U + k^2U = 0 \tag{6}$$

where

$$k^2 = \omega^2\epsilon\mu \tag{7}$$

and Δ is the Laplacian in spherical coordinates.

The field components can be obtained from equations (2) and (3). We find that

$$H_r = \left(k^2 + \frac{\partial^2}{\partial r^2}\right)(rU), \qquad H_\theta = \frac{1}{r}\frac{\partial^2}{\partial r\,\partial\theta}(rU),$$

(8)

$$E_\phi = -i\omega\mu\frac{\partial U}{\partial\theta}, \qquad E_r = E_\theta = H_\phi = 0.$$

The equations in (8) represent the field of a *magnetic dipole.* There are similar expressions for the field due to an electric dipole. The modifications of the theory for this case are detailed in Section 12.

The problem of the propagation of radio waves around the earth now has been reduced to that of solving the scalar wave equation (6). We know that

$$k^2(r) = \text{constant} = \omega^2\epsilon_1\mu, \qquad r < a$$

$$k^2(r) = \omega^2\epsilon(r)\mu, \qquad r > a$$

and that, at $r = a$, U and $\partial U/\partial r$ must be continuous because the tangential components of the field given by (8) must be continuous. We know also that U must be regular everywhere except at the dipole, that is, at $r = b$, $\theta = 0$. Finally, we know that the dipole is sending out radiation so that all the energy is going to infinity. This is expressed mathematically by the Sommerfeld radiation condition. Before writing it down, we remark that since the medium must eventually become free space, the value of ϵ will approach that of free space which we designate by ϵ_0. The Sommerfeld radiation condition may now be expressed as

$$\lim_{r\to\infty} r\left|\frac{\partial U}{\partial r} - ik_0U\right| = 0 \tag{9}$$

where

$$k_0^2 = \omega^2\epsilon_0\mu. \tag{10}$$

This condition means that, at infinity, U behaves like $\exp\{ik_0r\}/r$ so that we are dealing with an outgoing wave.

3. *The Series Solution for $U(r, \theta)$*

The fact that $U(r, \theta)$ possesses a dipole singularity at $r = b$, $\theta = 0$ while otherwise it satisfies the wave equation (6) can be expressed very concisely by using the Dirac δ-function. We write

$$\Delta U + k^2(r)U = -\frac{\delta(r-b)\,\delta(\theta)}{2\pi r^2\sin\theta} \tag{11}$$

where the denominator $2\pi r^2 \sin\theta$ has been introduced to normalize the solution. It is essentially the Jacobian of the transformation from rectangular to spherical coordinates.

Since $U(r, \theta)$ is a regular function of (r, θ) except at $r = b$, $\theta = 0$, we may assume

$$U = \sum_{n=0}^{\infty} \frac{2n+1}{2} u_n(r) P_n(\cos\theta). \tag{12}$$

From (12) by using the well known orthogonality properties of the Legendre polynomials we find that

$$u_n(r) = \int_0^{\pi} U(r, \theta) P_n(\cos\theta) \sin\theta d\theta. \tag{13}$$

Now, multiply (11) by P_n $(\cos\theta) \sin\theta d\theta$ and integrate between 0 and π. The equation transforms into the following ordinary differential equation:

$$\frac{1}{r}\frac{d^2(ru_n)}{dr^2} + \left[k^2(r) - \frac{n(n+1)}{r^2}\right]u_n = -\frac{\delta(r-b)}{2\pi r^2} \tag{14}$$

or

$$\frac{d^2(ru_n)}{dr^2} + \left[k^2(r) - \frac{n(n+1)}{r^2}\right](ru_n) = -\frac{\delta(r-b)}{2\pi r}.$$

Let us investigate the meaning of (14). If we integrate it first from $r = r_0$ to $r = b - h$ and then from $r = r_0$ to $r = b + h$, h a small quantity, we find that $d(ru_n)/dr$ has a jump of magnitude $-1/2\pi b$ as r goes from $b - h$ to $b + h$, while ru_n remains continuous. Note that the point $r = a$ causes no trouble because, there, just as for every value of $r \neq b$, the solution ru_n and its derivative must remain continuous. (These conditions are, of course, the same as those which have already been imposed on $U(r, \theta)$.) The other conditions on U, that it be regular at $r = 0$ and that it satisfy the Sommerfeld radiation condition (9), imply that u_n is regular at $r = 0$ and that

$$\lim_{r\to\infty} \left| \frac{d(ru_n)}{dr} - ik_0 ru_n \right| = 0. \tag{15}$$

In order to solve (14) we must know the solutions of the homogeneous equation

$$\frac{d^2(rv)}{dr^2} + \left[k^2(r) - \frac{n(n+1)}{r^2}\right](rv) = 0. \tag{16}$$

We shall study this equation, not only for integral values of n which are the only ones needed to solve (14), but also for complex values of n which will be needed later. In case $k^2(r)$ is a constant, k_0^2 say, the solutions of (16) are arbitrary

linear combinations of $r^{1/2}J_{n+\frac{1}{2}}(k_0r)$ and $r^{1/2}H^{(1)}_{n+\frac{1}{2}}(k_0r)$. The function $r^{1/2}J_{n+\frac{1}{2}}(k_0r)$ is regular at the origin if real part of $n > -1$, while the function $r^{1/2}H^{(1)}_{n+\frac{1}{2}}(k_0r)$ satisfies the radiation condition. If k^2 is not a constant, the solutions of (16) will behave approximately like the Bessel functions but their exact behavior will depend on the way k^2 varies with r.

We shall introduce three particular solutions of (16). Let $rv_n(r)$ be that solution which is regular at $r = 0$, while $rw^{(1)}_n(r)$ is that solution which for large r is asymptotically exp $\{ik_0r\}$, and $rw^{(2)}_n(r)$ is that solution which for large r is asymptotically exp $\{-ik_0r\}$. Note that the function $rw^{(1)}_n(r)$ satisfies the radiation condition (15). For simplicity of writing, we shall write it as $rw_n(r)$ when there is no possibility of confusion.

Since the differential equation (16) is unchanged when n is replaced by $-n - 1$, we see that

$$w^{(1)}_n(r) = w^{(1)}_{-n-1}(r), \qquad w^{(2)}_n(r) = w^{(2)}_{-n-1}(r).$$

This relation is not true for $v_n(r)$ because the condition of regularity at the origin will imply that $v_n(r)$ behaves like r^n, $n > 0$, at the origin and this condition is changed when n is replaced by $-n - 1$. It should also be noticed that in the case where k^2 has a constant value k_0^2 throughout, the function $rv_n(r)$ may be taken as

$$r^{1/2}J_{n+\frac{1}{2}}(k_0r);$$

on the other hand, because of the behavior at infinity, we must renormalize $rw^{(1)}_n(r)$ as follows:

$$rw^{(1)}_n(r) = \left(\frac{\pi}{2}\right)^{1/2} \exp\{+i\pi/4\} \exp\{+i(2n + 1)\pi/2\}(k_0r)^{1/2}H^{(1)}_{n+1/2}(k_0r)$$

and

$$rw^{(2)}_n(r) = \left(\frac{\pi}{2}\right)^{1/2} \exp\{-i\pi/4\} \exp\{-i(2n + 1)\pi/2\}(k_0r)^{1/2}H^{(2)}_{n+1/2}(k_0r).$$

We shall now prove the well-known fact that the Wronskian of any two solutions of (16) is a constant independent of r. Consider, for example, the two solutions $rv_n(r)$ and $rw_n(r)$. We have

$$\frac{d^2}{dr^2}(rv_n) + \left[k^2 - \frac{n(n+1)}{r^2}\right](rv_n) = 0$$

$$\frac{d^2}{dr^2}(rw_n) + \left[k^2 - \frac{n(n+1)}{r^2}\right](rw_n) = 0.$$

Multiply the first equation by rw_n, the second by rv_n and subtract. We get

$$(rw_n)\frac{d^2}{dr^2}(rv_n) - (rv_n)\frac{d^2}{dr^2}(rw_n) = 0.$$

Integrate this equation from $r = r_1$ to $r = r_2$, and we have

$$\left[rw_n(r) \frac{d}{dr}(rv_n) - rv_n(r) \frac{d}{dr}(rw_n) \right]_{r=r_1} = \left[rw_n(r) \frac{d}{dr}(rv_n) - rv_n(r) \frac{d}{dr}(rw_n) \right]_{r=r_2}$$

so that the Wronskian

$$W_n = rw_n(r) \frac{d}{dr}(rv_n) - rv_n(r) \frac{d}{dr}(rw_n) \tag{17}$$

is independent of the value of r but depends only on the value of n.

Now that we have some knowledge of the solutions of (16), we can prove that the solution of (14) is given by the following formulas:

$$\begin{aligned} ru_n(r) &= \frac{rv_n(r)w_n(b)}{2\pi W_n}, \qquad r < b \\ &= \frac{rv_n(b)w_n(r)}{2\pi W_n}, \qquad r > b. \end{aligned} \tag{18}$$

First, it is clear that both expressions in (18) satisfy (14) for $r \neq b$. Second, at $r = b$, the function ru_n is continuous and its derivative has a jump of magnitude $-1/2\pi b$. Third, from the definition of the functions $v_n(r)$ and $w_n(r)$, the function u_n is regular at $r = 0$ and satisfies the radiation condition (15). This shows that $ru_n(r)$ really is a solution of (14) satisfying all the other conditions.

Finally, we can obtain a representation for $U(r, \theta)$ from equation (12). We have

$$\begin{aligned} U(r, \theta) &= \frac{1}{2\pi} \sum_0^\infty \frac{v_n(r)w_n(b)}{W_n} \frac{2n+1}{2} P_n(\cos\theta), \qquad r < b \\ &= \frac{1}{2\pi} \sum_0^\infty \frac{v_n(b)w_n(r)}{W_n} \frac{2n+1}{2} P_n(\cos\theta), \qquad r > b. \end{aligned} \tag{19}$$

This is the solution of the problem in its most general form for an arbitrary variation of the dielectric constant. Special examples of this are well known for the cases in which the ordinary differential equation (16) can be solved in terms of known functions; for example, Rydbeck [3] has treated the problem where $\epsilon(r)$ is assumed to vary parabolically. It is to be noted that this solution (19) can be used for computation only if ka is small compared to unity. In the following sections this solution will be transformed so that it can be readily computed for large values of ka.

It is interesting to consider some special cases of (19). Suppose, first, that the medium is uniform and that there is no earth so that $k^2 = k_0^2$, a constant, then using the standard notation for the spherical Bessel function [13]

$$v_n(r) = j_n(k_0 r), \qquad w_n(r) = i^{n+1} h_n^{(1)}(k_0 r), \qquad W_n = i^n/k_0$$

formula (19) becomes

$$U(r,\ \theta) = \frac{\exp\{ik_0R\}}{4\pi R} = +\frac{k_0 i}{4\pi}\sum_0^\infty (2n+1)j_n(k_0r)h_n^{(1)}(k_0b)P_n(\cos\theta), \quad r < b$$

(20)

$$= +\frac{k_0 i}{4\pi}\sum_0^\infty (2n+1)j_n(k_0b)h_n^{(1)}(k_0r)P_n(\cos\theta), \quad r > b,$$

the classical formula for the dipole [13], where R is the distance from the point with coordinates $(r,\ \theta)$ to the source point $(b,\ 0)$. Suppose now that the earth is present so that for $r < a$, $k^2 = k_1^2$, $r > a$, $k^2 = k_0^2$; then

$$v_n(r) = j_n(k_1r), \qquad r < a$$

$$v_n(r) = \gamma_n j_n(k_0r) + \delta_n h_n^{(1)}(k_0r), \qquad r > a$$

$$w_n(r) = i^{n+1}h_n^{(1)}(k_0r), \qquad r > a$$

where γ_n, δ_n are to be determined by the fact that $u_n(r)$ and $(d/dr)\ u_n(r)$ are continuous at $r = a$. From the definition (17) of the Wronskian, it follows that

$$W_n = i^{n+1}b^2\gamma_n k_0[h_n^{(1)}(k_0b)j_n'(k_0b) - h_n^{(1)\prime}(k_0b)j_n(k_0b)]$$

$$= i^n\gamma_n/k_0\ .$$

γ is found from the following equations:

$$j_n(k_1a) = \gamma_n j_n(k_0a) + \delta_n h_n^{(1)}(k_0a)$$

$$k_1j_n'(k_1a) = \gamma_n k_0 j_n'(k_0a) + \delta_n k_0 h_n^{(1)\prime}(k_0a).$$

We have

$$\gamma_n = \frac{k_1j_n'(k_1a)h_n^{(1)}(k_0a) - k_0j_n(k_1a)h_n^{(1)}(k_0a)}{h_n^{(1)}(k_0a)j_n'(k_0a) - h_n^{(1)\prime}(k_0a)j_n(k_0a)}$$

so that

(21) $$W_n = i^{n+1}k_0a^2[k_1j_n'(k_1a)h_n^{(1)}(k_0a) - k_0j_n(k_1a)h_n^{(1)}(k_0a)].$$

Finally,

$$U(r,\ \theta) = \frac{1}{4\pi}\sum_0^\infty (2n+1)\frac{j_n(k_1r)h_n^{(1)}(k_0b)i^{n+1}P_n(\cos\theta)}{W_n}, \qquad r < a$$

$$= \frac{1}{4\pi}\sum_0^\infty (2n+1)\frac{\gamma_n j_n(k_0r) + \delta_n h_n^{(1)}(k_0r)}{W_n}\, i^{n+1}h_n^{(1)}(k_0b)P_n(\cos\theta),$$

(22) $$a < r < b$$

$$= \frac{1}{4\pi} \sum_0^\infty (2n+1) \frac{\gamma_n j_n(k_0 b) + \delta_n h_n^{(1)}(k_0 b)}{W_n} i^{n+1} h_n^{(1)}(k_0 r) P_n(\cos\theta),$$

$$r > b.$$

Notice that $\gamma_n/W_n = k_0 i^{-n}$ so that, by using (20), (22) can be written as the sum of two terms:

$$U(r,\theta) = \frac{\exp\{ik_0 R\}}{4\pi R} \tag{23}$$
$$+ \frac{1}{4\pi} \sum (2n+1) \frac{\delta_n h_n^{(1)}(k_0 b)}{W_n} i^{n+1} h_n^{(1)}(k_0 r) P_n(\cos\theta), \qquad r < b.$$

The first term is due to the dipole while the second term is due to the scattering of the dipole field by the earth.

In case the earth is assumed to be a perfect conductor so that $k_1 = \infty$, $j_n(k_1 a) = 0$ and, from the continuity condition, $\delta_n = -\gamma_n j_n(k_0 a)/h_n^{(1)}(k_0 a)$. We find then that

$$U(r,\theta) = \frac{ik_0}{4\pi} \sum (2n+1)[j_n(k_0 r) h_n^{(1)}(k_0 a) - h_n^{(1)}(k_0 r) j_n(k_0 a)]$$
$$\cdot \frac{h_n^{(1)}(k_0 b)}{h_n^{(1)}(k_0 a)} P_n(\cos\theta), \qquad a < r < b$$

(24)

$$= \frac{ik_0}{4\pi} \sum (2n+1)[j_n(k_0 b) h_n^{(1)}(k_0 a) - h_n^{(1)}(k_0 b) j_n(k_0 a)]$$
$$\cdot \frac{h_n^{(1)}(k_0 r)}{h_n^{(1)}(k_0 a)} P_n(\cos\theta), \qquad r > b.$$

4. *Contour Integral Representation for $U(r, \theta)$*

Following a method due to Watson [1], we shall show that the series (19) can be expressed as a contour integral. Consider the series

$$S = \sum_0^\infty \frac{2n+1}{2} a_n P_n(\cos\theta) \tag{25}$$

where a_n is assumed to be an analytic function of n. We shall show that S can be represented by the following integral:

$$I = \frac{1}{2\pi i} \int_C \frac{(t+\frac{1}{2}) a_t P_t(-\cos\theta)}{\sin \pi t} dt \tag{26}$$

where $C = C_1 + C_2$ is a contour that starts at $\infty - i\delta$ in the t-plane, goes below the real axis to $t = -1/2$ and then above the real axis to $\infty + i\delta$.

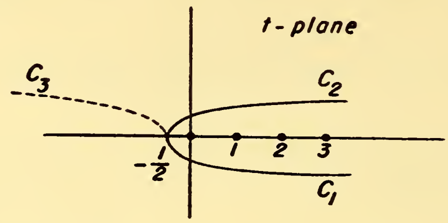


FIGURE 2

Since a_t and P_t $(-\cos\theta)$ are analytic functions of t, the only singularitie, of the integrand are poles at those values of t inside C for which $\sin \pi t = 0$s that is, $t = 0, 1, 2, \cdots$. Since the contour C is described in the clockwise direction, the integral I is equal to the negative sum of the residues at the poles, so that

$$(27) \qquad I = -\sum_0^\infty \frac{(n + \frac{1}{2})a_n P_n(-\cos\theta)}{\pi \cos \pi n} = -\frac{1}{\pi}\sum_0^\infty \left(n + \frac{1}{2}\right)a_n P_n(\cos\theta)$$

because $P_n(-\cos\theta) = (-1)^n P_n(\cos\theta)$.

Formula (27) shows that $S = -\pi I$. If this procedure is applied to (19), we find that

$$(28) \qquad U(r, \theta) = -\frac{1}{4\pi i}\int_C \frac{(t + \frac{1}{2})v_t(b)w_t(r)}{W_t}\frac{P_t(-\cos\theta)}{\sin \pi t}\,dt$$

where C is the contour specified above. On C_1, the part of C below the real axis, replace t by $-t - 1$, then C_1 is transformed into C_3 and we have

$$(29) \qquad \begin{aligned} U(r, \theta) = &-\frac{1}{4\pi i}\int_{C_3} \frac{(t + \frac{1}{2})v_{-t-1}(b)w_{-t-1}(r)}{W_{-t-1}}\frac{P_{-t-1}(-\cos\theta)}{\sin \pi t}\,dt \\ &-\frac{1}{4\pi i}\int_{C_2} \frac{(t + \frac{1}{2})v_t(b)w_t(r)}{W_t}\frac{P_t(-\cos\theta)}{\sin \pi t}\,dt. \end{aligned}$$

We shall rotate C_2 and C_3 so that they go around the upper half of the line, imaginary part of $t = -1/2$. To justify this, it is necessary to show that the integrals over the infinite quarter-circles in the first and second quadrants go to zero. This is discussed in the appendix.

When the contour is rotated in the desired manner, it may pass across some poles of the integrand. We find then that $U(r, \theta)$ as given by (29) can be written as the sum of residues at the poles plus two integrals over the upper half of the line $\mathscr{I}m\ (t) = -1/2$. Put $t = -1/2 + i\tau$ and use the fact [14] that

$$(30) \qquad P_t(-\cos\theta) = P_{-t-1}(-\cos\theta),$$

we obtain

$$U(r,\theta) = \sum \text{residues} + \frac{1}{4\pi i}\int_0^\infty \frac{\tau P_{i\tau-\frac{1}{2}}(-\cos\theta)}{-\cosh \pi\tau}\left[\frac{v_{i\tau-\frac{1}{2}}(b)w_{i\tau-\frac{1}{2}}(r)}{W_{i\tau-\frac{1}{2}}} - \frac{v_{-i\tau-\frac{1}{2}}(b)w_{-i\tau-\frac{1}{2}}(r)}{W_{-i\tau-\frac{1}{2}}}\right]d\tau \tag{31}$$

From (28) it is clear that the poles of the integrand will be at the points

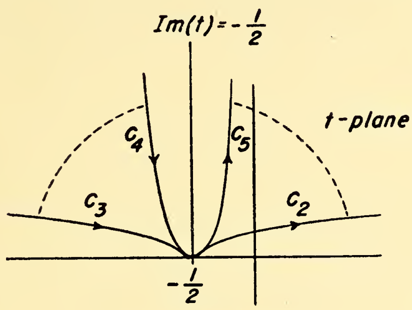


FIGURE 3

where $W_t = 0$. Call these points $t_1, t_2, \cdots$; then the residue at any pole, t_j say, is

$$\frac{(t_j + \frac{1}{2})v_{t_j}(b)w_{t_j}(r)P_{t_j}(-\cos\theta)}{W'_{t_j}\sin \pi t_j}. \tag{32}$$

Here

$$W'_{t_j} = \frac{d}{dt}W_t\bigg|_{t=t_j}. \tag{33}$$

We thus have the final result:

$$U(r,\theta) = \sum_j \frac{(t_j + \frac{1}{2})v_{t_j}(b)w_{t_j}(r)P_{t_j}(-\cos\theta)}{W'_{t_j}\sin \pi t_j} + \frac{1}{4\pi i}\int_0^\infty \frac{\tau P_{i\tau-\frac{1}{2}}(-\cos\theta)}{-\cosh\pi\tau}\left[\frac{v_{i\tau-\frac{1}{2}}(b)w_{i\tau-\frac{1}{2}}(r)}{W_{i\tau-\frac{1}{2}}} - \frac{v_{-i\tau-\frac{1}{2}}(b)w_{-i\tau-\frac{1}{2}}(r)}{W_{-i\tau-\frac{1}{2}}}\right]d\tau.$$

5. *The Eigenvalue Problem*

The residue in (32) can be more usefully interpreted in another way. Since the Wronskian, $W_{t_j} = 0$, the two functions $v_{t_j}(r)$ and $w_{t_j}(r)$ are not independent, that is, there exist constants α_j such that $v_{t_j}(r) = \alpha_j w_{t_j}(r)$. This equation shows that the function $w_{t_j}(r)$, besides satisfying the radiation condition at infinity, also satisfies the condition of regularity at $r = 0$. In other

words, t_i is an eigenvalue and $w_{t_i}(r)$ is an eigenfunction of the following problem:

Find the values of t for which $v(r)$, the solution of the following differential equation:

$$\frac{d^2(rv)}{dr^2} + \left[k^2 - \frac{t(t+1)}{r^2}\right](rv) = 0, \tag{34}$$

is regular at $r = 0$ and satisfies the radiation condition (15) at infinity, that is,

$$\lim_{r\to\infty} \left| \frac{d(rv)}{dr} - ik_0 rv \right| = 0.$$

Put

$$\beta_i = \int_0^\infty w_{t_i}(r)^2 \, dr \tag{35}$$

so that $w_{t_i}(r)/\beta_i^{1/2}$ is a normalized eigenfunction of (34); then the residue given by (32) will be shown to be equal to

$$\frac{w_{t_i}(b) w_{t_i}(r) P_{t_i}(-\cos\theta)}{2\beta_i \sin \pi t_i}. \tag{36}$$

This result will be obtained from a consideration of the following two equations:

$$\frac{d^2(rv_t)}{dr^2} + \left[k^2 - \frac{t(t+1)}{r^2}\right](rv_t) = 0$$

$$\frac{d^2(rw_{t_i})}{dr^2} + \left[k^2 - \frac{t_i(t_i+1)}{r^2}\right](rw_{t_i}) = 0.$$

Here t has any value and is not an eigenvalue. Multiply the first equation by rw_{t_i}, the second by rv_t subtract and integrate with respect to r from 0 to ∞. We obtain

$$(t - t_i)(t + t_i + 1) \int_0^\infty v_t w_{t_i} \, dr = rv_t \frac{d}{dr}(rw_{t_i}) - rw_{t_i} \frac{d}{dr}(rv_t) \Big|_0^\infty. \tag{37}$$

Notice the term at zero drops out because by definition v_t and w_{t_i} are regular at $r = 0$.

Consider the following determinant:

$$\begin{vmatrix} rv_t & rw_t & rw_{t_i} \\ rv_t & rw_t & rw_{t_i} \\ \frac{d}{dr}(rv_t) & \frac{d}{dr}(rw_t) & \frac{d}{dr}(rw_{t_i}) \end{vmatrix}.$$

It is obviously zero since the elements of the first two rows are identical. When it is expanded in terms of the elements of the first two rows, we have

$$rv_t\left[rw_t\frac{d}{dr}(rw_{t_i}) - rw_{t_i}\frac{d}{dr}(rw_t)\right] + rw_{t_i}\left[rv_t\frac{d}{dr}(rw_t) - rw_t\frac{d}{dr}(rv_t)\right]$$

$$- rw_t\left[rw_{t_i}\frac{d}{dr}(rv_t) - rv_t\frac{d}{dr}(rw_{t_i})\right] = 0.$$

Now, as r approaches infinity, the first bracket approaches zero since rw_t and rw_{t_i} both satisfy the same condition at infinity. This shows that

$$\left[rv_t\frac{d}{dr}(rw_{t_i}) - rw_{t_i}\frac{d}{dr}(rv_t)\right] = \frac{w_{t_i}\left[rw_t\frac{d}{dr}(rv_t) - rv_t\frac{d}{dr}(rw_t)\right]}{w_t}$$

so that (37) can be written as follows:

$$\int_0^\infty v_t w_{t_i}\, dr = \frac{w_{t_i}(r)}{w_t(r)}\frac{W_t}{(t - t_i)(t + t_i + 1)}\,.$$

As t approaches t_i , w_t approaches w_{t_i} and v_t approaches $\alpha_i w_{t_i}$ while the right side of the above equation becomes indeterminate. If the right side is evaluated by L'Hôpital's rule, we find that

$$\alpha_i\int_0^\infty w_{t_i}(r)^2\, dr = \frac{W_t'}{2t_i + 1}$$

so that

$$\beta_i = \frac{W_t'}{(2t_i + 1)\alpha_i}\,;$$

using this in (32) we get (36).

When (36) is substituted in (31) we get finally the result that

$$U(r,\theta) = \sum\frac{w_{t_i}(b)w_{t_i}(r)}{2\beta_i}\frac{P_{t_i}(-\cos\theta)}{\sin\pi t_i} + I \tag{38}$$

where I is the following integral:

$$I = \frac{1}{4\pi i}\int_0^\infty\frac{\tau P_{i\tau-\frac{1}{2}}(-\cos\theta)}{\cosh\pi\tau}\left[\frac{v_{-i\tau-\frac{1}{2}}(b)w_{-i\tau-\frac{1}{2}}(r)}{W_{-i\tau-\frac{1}{2}}} - \frac{v_{i\tau-\frac{1}{2}}(b)w_{i\tau-\frac{1}{2}}(r)}{W_{i\tau-\frac{1}{2}}}\right]d\tau.$$

In the appendix it will be shown that, if the earth is assumed to be a perfect conductor, then I vanishes identically. If the earth is not assumed to be a perfect conductor, I will not be zero. However, it will be shown that I is small compared to the first term of the series in (38) as long as the dielectric constant of the earth is large compared to the dielectric constant of the atmosphere.

6. *A Simpler Form for the Hertz Potential*

So far, we have not yet made use of the fact that the dielectric constant of the earth is a constant, nor of the fact that, for the wavelengths we are interested in, ka is very large of the order of magnitude of 10^6. We shall use these facts to simplify the expansion given in (38).

It will be seen in the later sections that the zeros of the Wronskian, t_j , will all be very large, of the order of magnitude of ka. We are therefore justified in using an asymptotic formula for the Legendre functions that appear in (38). We have [14]

$$P_\nu(\cos\theta) \sim \left(\frac{2}{\pi(\nu+1)\sin\theta}\right)^{1/2} \cos\left\{\left(\nu+\frac{1}{2}\right)\theta - \frac{\pi}{4}\right\}$$

for large values of ν, so that

$$\text{(40)} \qquad \begin{aligned} \frac{P_t(-\cos\theta)}{\sin \pi t_j} &= \frac{P_{t_j}(\cos(\pi-\theta))}{\sin \pi t_j} \\ &\sim -\left(\frac{2i}{\pi(t_j+1)\sin\theta}\right)^{1/2} \exp\left\{i\left(t_j+\frac{1}{2}\right)\theta\right\}. \end{aligned}$$

Next, since for $r < a$, $k^2 = k_1^2 =$ constant, we know that the solutions of (34) are

$$rv = r^{1/2}J_{t+\frac{1}{2}}(k_1 r) \qquad \text{or} \qquad r^{1/2}H^{(1)}_{t+\frac{1}{2}}(k_1 r), \qquad r < a.$$

The function $v_t(r)$ was defined to be that solution of (34) which was regular at $r = 0$. It must, therefore, be proportional to the function $J_{t+\frac{1}{2}}(k_1 r)$ for $r < a$. Since the solution of (34) and its first derivative must be continuous at $r = a$, this requires that

$$\text{(41)} \qquad \left.\frac{(d/dr)rv_t(r)}{rv_t(r)}\right|_{r=a} = \left.\frac{(d/dr)(r^{1/2}J_{t+\frac{1}{2}}(k_1 r))}{r^{1/2}J_{t+\frac{1}{2}}(k_1 r)}\right|_{r=a}.$$

Again, the right side can be simplified by using the asymptotic formulas. Put

$$g(r) = \frac{(d/dr)(r^{1/2}J_{t+\frac{1}{2}}(k_1 r))}{r^{1/2}J_{t+\frac{1}{2}}(k_1 r)}$$

so that

$$r^{1/2}J_{t+\frac{1}{2}}(k_1 r) = C\exp\left\{\int g(r)\,dr\right\}$$

where C is a suitable constant. Substituting this expression in the differential equation for $r^{1/2}J_{t+\frac{1}{2}}(k_1 r)$, that is,

$$\frac{d^2(rv)}{dr^2} + \left[k_1^2 - \frac{t(t+1)}{r^2}\right](rv) = 0,$$

we get

$$\frac{dg}{dr} + g^2 + k_1^2 - \frac{t(t+1)}{e^2} = 0. \tag{42}$$

Since $k_1^2 = \omega^2 \epsilon_1 \mu$, where ϵ_1, the dielectric constant of the earth, is very large, we may neglect dg/dr in (42) and get the approximate formula:

$$g(r) = \pm i \left[k_1^2 - \frac{t(t+1)}{r^2} \right]^{1/2}.$$

To determine whether it is a plus or minus sign, we notice that ϵ_1 and so, also k_1, has a large positive imaginary part so that, by the usual asymptotic formula,

$$r^{1/2} J_{t+\frac{1}{2}}(k_1 r) \sim \left(\frac{2}{\pi}\right)^{1/2} \cos \left\{ k_1 r - \frac{\pi}{2}\left(t + \frac{1}{2}\right) - \frac{\pi}{4} \right\} \sim C_2 \exp \{-ik_1 r\}$$

where C_2 is independent of r. Then

$$\frac{(d/dr)(r^{1/2} J_{t+\frac{1}{2}}(k_1 r))}{r^{1/2} J_{t+\frac{1}{2}}(k_1 r)} \sim -ik_1$$

and so $g(r)$, which is a more accurate expression for the ratio of the derivative to the function must have the minus sign, that is,

$$g(r) = -i[k_1^2 - t(t+1)/r^2]^{1/2}.$$

We shall see later that $t \sim ak(a)$. Because of this, $g(a) = -i(k_1^2 - k^2)^{1/2}$, a quantity independent of t. Now (41) can be written as follows:

$$\left. \frac{(d/dr) r v_t(r)}{r v_t(r)} \right|_{r=a} = -i(k_1^2 - k^2)^{1/2}. \tag{43}$$

This result holds for all values of t, if t is approximately $ak(a)$.

Consider the eigenvalues t_i for which the functions $v_t(r)$ and $w_t(r)$ coincide. Then (43) gives us the following eigenvalue equation:

$$\left. \frac{(d/dr) r w_{t_i}(r)}{r w_{t_i}(r)} \right|_{r=a} = -i(k_1^2 - k^2)^{1/2}. \tag{44}$$

We shall discuss its solution in a later section.

There is one more quantity in the expansion (38) that will be simplified. It is the quantity β_i or essentially W_t'. Since the Wronskian is constant for all values of r, we have

$$W_t = r w_t \frac{d}{dr}(r v_t) \Big|_{r=a} - r v_t \frac{d}{dr}(r w_t) \Big|_{r=a}$$

and $$W_t' = \frac{\partial (r w_t)}{\partial t} \frac{d}{dr}(r v_t) + r w_t \frac{\partial}{\partial t}\frac{d}{dr}(r v_t) - \frac{\partial (r v_t)}{\partial t}\frac{d}{dr}(r w_t) - r v_t \frac{\partial}{\partial t}\frac{\partial}{\partial r}(r w_t),$$

all evaluated at $r = a$. Now, since the right side of (43) is independent of t, we have

$$\frac{\partial}{\partial t} \frac{(d/dr)(rv_t)}{rv_t}\bigg|_{r=a} = 0$$

or

$$rv_t \frac{\partial}{\partial t} \frac{d}{dr}(rv_t)\bigg|_a - \frac{\partial(rv_t)}{\partial t} \frac{d}{dr}(rv_t)\bigg|_a = 0.$$

When $t = t_i$, an eigenvalue, the function $v_{t_i}(r)$ is identical with $w_{t_i}(r)$ and then

$$rw_{t_i} \frac{\partial}{\partial t} \frac{d}{dr}(rv_{t_i}) - \frac{\partial(rv_{t_i})}{\partial t} \frac{d}{dr}(rv_{t_i}) = 0.$$

Using this result in the formula for W'_t, we find that

$$W'_{t_i} = \frac{\partial(rw_{t_i})}{\partial t} \frac{d}{dr}(rw_{t_i})\bigg|_a - rw_{t_i} \frac{\partial}{\partial t} \frac{d}{dr}(rw_{t_i})\bigg|_a .$$

If $w_{t_i}(a) \neq 0$, which by (44) will always be the case if $k_1 \neq \infty$, that is, the earth is not a perfect conductor, then we may write the above formula as follows:

$$W'_{t_i} = (rw_{t_i})^2 \frac{\partial}{\partial t} \frac{(d/dr)(rw_{t_i})}{rw_{t_i}}\bigg|_a = (aw_{t_i})^2 M, \tag{45}$$

where M is defined by the last equation.

Combining the results we have obtained in (45), (44), (42) and (38), we get, finally, the desired expression:

$$\begin{aligned} U(r, \theta) &= -\sum_i \left[\frac{(2t_i + 1)}{\pi \sin \theta}\right]^{1/2} \exp\left\{i\left(t_i + \frac{1}{2}\right)\theta\right\} \frac{w_{t_i}(b) w_{t_i}(r)}{W'_{t_i}} \\ &= \frac{-1}{br} \sum_i \left[\frac{(2t_i + 1)i}{\pi \sin \theta}\right]^{1/2} \frac{\exp\{i(t_i + \frac{1}{2})\theta\}}{M} \frac{bw_{t_i}(b)}{aw_{t_i}(a)} \frac{rw_{t_i}(r)}{aw_{t_i}(a)} . \end{aligned} \tag{46}$$

Here, $bw_{t_i}(b)/aw_{t_i}(a)$ is the height gain factor for the transmitter and $rw_{t_i}(r)/aw_{t_i}(a)$ is the height gain factor for the receiver.

7. *The Langer Asymptotic Formulas*

The previous sections have shown that the problem of determining the Hertz vector $U(r, \theta)$ and thus the electromagnetic field depends upon the solution of the eigenvalue problem for the differential equation (34). In general, this problem has to be solved by some approximation. In this section we shall discuss by a method due to Langer [9] the well-known W.K.B. approximation to the solutions of a second-order differential equation and in later sections we shall use these approximations to obtain the eigenvalues of (34):

(47) $$\frac{d^2(rv)}{dr^2} + \left[k(r)^2 - \frac{\lambda}{r^2}\right](rv) = 0.$$

Here, we have put λ for $t(t+1)$.

Consider a second order differential equation of the form:

(48) $$\frac{d^2s}{dr^2} + [\nu^2 f(r) + g(r)]s = 0$$

where ν is assumed to be a large constant and where $f(r) = (r - r_0)^\alpha f_0(r)$, α is a constant and $f_0(r_0) \neq 0$. Put

(49) $$r - r_0 = h(\tau), \qquad s = \sigma(h'(\tau))^{1/2}$$

where $h(\tau)$ is so chosen that

(50) $$h'(\tau)^2 f(r) = 1$$

and $\tau = 0$ for $r = r_0$. Under the transformations (49) equation (48) becomes

(51) $$\frac{d^2\sigma}{d\tau^2} + \left\{\nu^2 + \frac{\alpha(\alpha+4)}{4(\alpha+2)^2\tau^2} + \psi(\tau)\right\}\sigma = 0$$

where $\psi(\tau)$ is small for large values of ν. The solutions of (51) are, approximately, linear combinations of

(52) $$\tau^{1/2}H_\gamma^{(1)}(\nu\tau) \qquad \text{and} \qquad \tau^{1/2}H_\gamma^{(2)}(\nu\tau)$$

where $\gamma = (\alpha + 2)^{-1}$.

We shall see later that for our purposes only the cases $\alpha = 1$ and $\alpha = 2$ are of interest. The case $\alpha = 1$ corresponds to the situation where the atmosphere does not have a duct or a temperature inversion. The case $\alpha = 2$ will correspond to the situation where the atmosphere has a duct.

In case $\alpha = 1$, we rewrite equation (47) as follows:

$$\frac{d^2(rv)}{dr^2} + k_0^2a^2\left[\left(\frac{k(r)}{ak_0}\right)^2 - \frac{\lambda/a^2}{(k_0r)^2}\right](rv) = 0.$$

This equation is of the same form as (48) if we make k_0a correspond to ν and put

$$f(r) = \left(\frac{k(r)}{k_0a}\right)^2 - \frac{\lambda/a^2}{(k_0r)^2}, \qquad g(r) = 0.$$

To use the formulas (52) we must put from (50)

$$\frac{d\tau}{dr} = \frac{1}{h'(\tau)} = [f(r)]^{1/2} \qquad \text{or}$$

(53) $$\tau = \frac{1}{a}\int_{r_0}^{r}\left[\left(\frac{k(r)}{k_0}\right)^2 - \frac{\lambda}{(k_0r)^2}\right]^{1/2} dr.$$

Here, r_0 is taken to be a zero of $f(r)$ so that $r_0^2 k(r_0)^2 = \lambda$. If we change also the dependent variable from $v(r)$ to $\sigma(\tau)$ by the transformation indicated in (49):

$$rv = \sigma[h'(\tau)]^{1/2} = \sigma\left[\left(\frac{k(r)}{k_0 a}\right)^2 - \frac{\lambda/a^2}{(k_0 r)^2}\right]^{-1/4},$$

then, according to (52), we have

$$rw_t^{(1,2)} = \tau^{1/2}\left[\left(\frac{k(r)}{k_0 a}\right)^2 - \frac{\lambda/a^2}{(k_0 r)^2}\right]^{-1/4} H_{1/3}^{(1,2)}(k_0 a\tau).$$

In particular, the solution which satisfies the Sommerfeld radiation condition at infinity will be given by the formula:

$$rw_t = \tau^{1/2}\left[\left(\frac{k(r)}{k_0 a}\right)^2 - \frac{\lambda/a^2}{(k_0 r)^2}\right]^{-1/4} H_{1/3}^{(1)}(k_0 a\tau). \tag{54}$$

It should be noticed that the W.K.B. method and the Langer formulas are not valid if in (48) $f(r)$ is discontinuous. This means that formula (54) is not valid for $r < a$. However, we have already seen that the solution of (34) for $r < a$ behaves like the Bessel function $J_{t+\frac{1}{2}}(k_1 r)$. Using the approximate formula (54) and neglecting terms of order $(ka)^{-1}$, we obtain a simpler form for the eigenvalue equation (44). It becomes

$$\left[\left(\frac{k(a)}{k_0}\right)^2 - \frac{\lambda}{(k_0 a)^2}\right]^{1/2} \frac{H_{1/3}^{(1)}(k_0 a\tau_a)'}{H_{1/3}^{(1)}(k_0 a\tau_a)} = -i\frac{k_1}{k_0}\left[1 - \left(\frac{k(a)}{k_1}\right)^2\right]^{1/2}. \tag{55}$$

Here τ_a is the value of τ corresponding to $r = a$. Note that, if the earth is assumed to be a perfect conductor so that $k_1 = \infty$, then equation (55) reduces to the following simple form:

$$H_{1/3}^{(1)}(k_0 a\tau_a) = 0. \tag{56}$$

Since the ratio

$$\frac{dH_{1/3}^{(1)}(x)}{dx} \Big/ H_{1/3}^{(1)}(x)$$

will appear so frequently in our analysis, we shall use for it the simpler notation $Z(x)$. Equation (55) can now be written as follows:

$$\left[\left(\frac{k(a)}{k_0}\right)^2 - \frac{\lambda}{(k_0 a)^2}\right]^{1/2} Z(k_0 a\tau_a) = \frac{-ik_1}{k_0}\left[1 - \left(\frac{k(a)}{k_1}\right)^2\right]^{1/2}.$$

Put

$$y = C_1 H_{1/3}^{(1)}(x) = \exp\left\{\int_\alpha^x Z(t)\,dt\right\}$$

where C_1 and α are suitably chosen constants. From the differential equation satisfied by y

$$y'' + \frac{1}{x}y' + \left(1 - \frac{1}{9x^2}\right)y = 0$$

we find that Z satisfies the Riccati equation:

$$Z' + Z^2 + \frac{1}{x} Z + 1 - \frac{1}{9x^2} = 0$$

so that for large x

(57) $$Z' \sim -1 - Z^2 - \frac{1}{x} Z.$$

8. *The Case of a Uniform Atmosphere*

We shall discuss first the case of a uniform atmosphere in order to more easily understand the case of a non-uniform atmosphere. The methods and results of this section are classical [11]. Suppose that the atmosphere is uniform so that $k^2 = k_0^2$ = constant and also that the earth is a perfect conductor. The solutions of

$$H_{1/3}^{(1)}(\xi) = 0$$

are

(58) $$\xi_0 = 2.38\, e^{-\pi i}, \qquad \xi_1 = 5.6\, e^{-\pi i}, \qquad \xi_2 = 8.65\, e^{-\pi i}, \text{ etc.}$$

The equation (56) now becomes the following:

(59) $$k_0 \int_{r_0}^{a} \left[1 - \frac{\lambda}{(k_0 r)^2}\right]^{1/2} dr = \xi_n = e^{-\pi i}\tau_n ,$$

where ξ_n is defined by the last equation.

Put

(60) $$\rho = k_0 r \lambda^{-1/2}, \qquad \rho_a = k_0 a \lambda^{-1/2}$$

in (59). Since $k_0 r_0 = \lambda^{1/2}$, it reduces to

$$\lambda^{1/2} \int_1^{\rho_a} (1 - \rho^{-2})^{1/2}\, d\rho = e^{-\pi i}\tau_n .$$

When we evaluate the integral we get, finally, the eigenvalue equation,

(61) $$(\rho_a^2 - 1)^{1/2} - \arctan(\rho_a^2 - 1)^{1/2} = e^{-\pi i}\tau_n \lambda^{-1/2}.$$

In this equation ρ_a is a function of λ defined by (60).

We shall find that $\lambda^{1/2}$ is of the order of magnitude of $(k_0 a)$ and therefore the right side of (60) will be very near to zero which implies that ρ_a will be close to one.

Put $\rho_a = 1 + \delta(k_0 a)^{-2/3}$ then (61) becomes $(2\delta)^{3/2}(k_0 \alpha)^{-1}/3 = e^{-\pi i}\tau_n\lambda^{-1/2} = e^{-\pi i}\tau_n(k_0 a)^{-1}$ or

(62) $$\delta = \tfrac{1}{2}(3\tau_n)^{2/3} e^{-2\pi i/3}.$$

Each value of τ_n determines a value of δ, then a value of ρ_a, and by (60) a value of λ. Since t is approximately $\lambda^{1/2}$, we finally get the well-known expression [11] for the eigenvalues:

(63) $$t_n = k_0a + \tfrac{1}{2}(k_0a)^{1/3}(3\tau_n)^{2/3}e^{\pi i/3}.$$

Note that (62) shows that δ is of the order of magnitude one and since $(k_0a)^{-2/3}$ is small, our assumption that ρ_a is close to one is justified. It is also important to notice that the integral below (60) can be easily evaluated approximately. Since ρ_a is very close to one, we may proceed as follows:

$$\int_1^{\rho_a} (1 - \rho^{-2})^{1/2}\, d\rho \sim \int_1^{\rho_a} (\rho^2 - 1)^{1/2}\, d\rho \sim 2^{1/2} \int_1^{\rho_a} (\rho - 1)^{1/2}\, d\rho$$

$$= \frac{1}{3}\, 2^{3/2}(\rho_a - 1)^{3/2} = \frac{1}{3}\,(2\delta)^{3/2}(k_0a)^{-1}$$

which is the same result that was obtained by an exact integration.

If we drop the assumption that the earth is a perfect conductor, we must solve the eigenvalue equation (55) instead of the equation $H_{1/3}^{(1)}(\xi) = 0$. Using the same approximations and notations as before, we find that

$$\left[1 - \frac{\lambda}{(k_0a)^2}\right]^{1/2} = \left[1 - \frac{1}{\rho_a^2}\right]^{1/2} \sim [2(\rho_a - 1)]^{1/2} = (2\delta)^{1/2}(k_0a)^{-1/3}$$

and that

$$k_0a\tau = k_0 \int_{r_0}^{a} \left[1 - \frac{\lambda}{(k_0r)^2}\right] dr$$

$$= \lambda^{1/2} \int_1^{\rho_a} (1 - \rho^{-2})^{1/2}\, d\rho \sim \frac{\lambda^{1/2}}{3}\,(k_0a)^{-1}(2\delta)^{3/2} \sim \frac{1}{3}\,(2\delta)^{3/2}$$

so that (55) becomes

(64) $$(k_0a)^{-1/3}(2\delta)^{1/2}Z\left\{\frac{(2\delta)^{3/2}}{3}\right\} = -i\,\frac{k_1}{k_0}\left[1 - \left(\frac{k(a)}{k_1}\right)^2\right]^{1/2} = \eta,$$

where η is defined by the last equation. Methods for solving this equation are discussed by Bremmer [11].

We shall now find a simple approximate formula for the quantity M which appears in (45) and then we can write down the complete formula for the Hertz potential as given by (46). From (45), (57) and (64) we find that

$$M = \frac{\partial}{\partial t}\,\frac{(d/dr)(rw_{t_i})}{rw_{t_i}}\bigg|_a = k_0(k_0a)^{-1/3}\,\frac{\partial}{\partial t}\left[(2\delta)^{1/2}Z\left\{\frac{(2\delta)^{3/2}}{3}\right\}\right]$$

$$= k_0(k_0a)^{-1/3}\,\frac{\partial\delta}{\partial t}\,\frac{\partial}{\partial\delta}\left[(2\delta)^{1/2}Z\left\{\frac{(2\delta)^{3/2}}{3}\right\}\right]$$

$$= k_0(k_0a)^{-1/3}\,\frac{\partial\delta}{\partial t}\left[\frac{1}{(2\delta)^{1/2}}\,Z + 2\delta Z'\right]$$

$$= k_0(k_0a)^{-1/3}\,\frac{\partial\delta}{\partial t}\left[\frac{1}{(2\delta)^{1/2}}\,Z - 2\delta\left\{1 + Z^2 + \frac{3}{(2\delta)^{3/2}}\,Z\right\}\right]$$

$$= -k_0(k_0a)^{-1/3}\frac{\partial\delta}{\partial t}\left[2\delta + 2\delta Z^2 + \frac{2}{(2\delta)^{1/2}}Z\right]$$

$$= -k_0(k_0a)^{-1/3}\frac{\partial\delta}{\partial t}\left[2\delta + \eta^2(k_0a)^{2/3} + \frac{2\eta(k_0a)^{1/3}}{2\delta}\right].$$

Now, by (60)

$$t(t+1) = \lambda = (k_0a)^2\rho_a^{-2} = (k_0a)^2[1 + \delta(k_0a)^{-2/3}]^{-2}$$

so that, approximately,

$$t = k_0a[1 + \delta(k_0a)^{-2/3}]^{-1} = k_0a - \delta(k_0a)^{1/3}$$

and then

$$\frac{\partial\delta}{\partial t} = -(k_0a)^{-1/3}.$$

Using this result in the expression for M given above, we find, finally, that

$$M = k_0(k_0a)^{-2/3}[2\delta + \eta^2(k_0a)^{2/3} + \eta(k_0a)^{1/3}\delta^{-1}], \tag{65}$$

Consider the case where both the transmitter and receiver are on the earth's surface so that $b = a$ and $r = a$. Substituting (65) and (63) in (46) we get

$$\begin{aligned} U(a, \theta) &= -\frac{\exp\{ik_0a\theta\}}{\pi a\theta}[2\pi i(k_0a)^{1/3}\theta]^{1/2} \\ &\quad \cdot \sum_i \frac{\exp\{i(k_0a)^{1/3}\theta(3\tau_i)^{2/3}(e^{\pi i/3})/4\}}{2\delta_i + \eta^2(k_0a)^{2/3} + \eta(k_0a)^{1/3}\delta^{-1}} e^{i\theta/2}. \end{aligned} \tag{66}$$

Here τ_i is the root of (64). Note that $a\theta = d$ is the distance from the transmitter to the receiver so that

$$U(a, \theta) = -\frac{\exp\{ik_0d\}}{\pi d}(2\pi i\chi)^{1/2}F(\chi)$$

where $\chi = (k_0a)^{1/3}\theta$

$$F(\chi) = \sum_i \frac{\exp\{\chi(3\tau_i)^{2/3}(e^{\pi i/3})/4\}}{2\delta_i + \eta^2 + \eta(k_0a)^{1/3}\delta^{-1}} e^{i\theta/2}.$$

As was pointed out before, these results are well known.

9. *The Case of a Non-uniform Atmosphere*

If the atmosphere is not uniform and is such that $r^2k(r)^2$ has no stationary point, that is, physically the atmosphere does not have a duct, then it can be treated similarly to the case of a uniform atmosphere. We shall see that the results of the preceding sections will be only slightly modified.

First, we find an approximate formula for τ as defined in (53). Since a will

be near to r_0 and $\lambda^{1/2}$ to $ak(a)$, the bracket in (53) may be approximated as follows. Put

$$k(a) = k_a\,, \qquad [rk(r)]'_{r=a} = \zeta, \qquad \zeta k_a k_0^{-2} = K^2.$$

Then

$$\left(\frac{k(r)}{k_0}\right)^2 - \frac{\lambda}{(k_0 r)^2} \sim \frac{r^2k(r)^2 - \lambda}{(k_0 a)^2} \sim \frac{2ak_a}{(k_0 a)^2}\,[rk(r) - \lambda^{1/2}]$$

(67)

$$\sim \frac{2ak_a}{(k_0 a)^2}\,\zeta(r - r_0) \sim 2\,\frac{k_a\zeta}{k_0^2}\left(\frac{r}{r_0} - 1\right) = 2K^2\left(\frac{r}{r_0} - 1\right).$$

Here, we have made use of the fact that since $r_0 k(r_0) = \lambda^{1/2}$, $rk(r) - \lambda^{1/2} = \zeta(r - r_0)$ approximately.

Now,

$$k_0 a\tau = 2^{1/2}k_0 K \int_{r_0}^{r} \left(\frac{r}{r_0} - 1\right)^{1/2} dr = 2^{3/2}Kk_0 a\left(\frac{r}{r_0} - 1\right)^{3/2}/3.$$

If we put, as before, $a = r_0[1 + \delta(Kk_0a)^{-2/3}]$ then $k_0 a\tau_a = (2\delta)^{3/2}/3$ and the eigenvalue equation (55) becomes

$$(Kk_0a)^{-1/3}(2\delta)^{1/2}Z\left\{\frac{(2\delta)^{3/2}}{3}\right\} = -i\,\frac{k_1}{Kk_0}\left[1 - \frac{k_a^2}{k_1^2}\right]^{1/2} = \bar{\eta}, \tag{68}$$

where $\bar{\eta}$ is defined by the last equation.

Now $t \sim \lambda^{1/2} = r_0 k(r_0) = ak_a + \zeta(r_0 - a)$, by Taylor's Theorem, but $r_0 - a = -\delta r_0(Kk_0a)^{-2/3} \sim -a\delta(Kk_0a)^{-1/3}$ and so

$$t = ak_a - \delta(Kk_0a)^{1/3}(Kk_0/k_a). \tag{69}$$

In most cases the term $ak'(a)$ will be negligible compared to k_a, so that $\zeta = k_a$ and $K = k_a/k_0$. Then

$$t = ak_a - \delta(k_a a)^{1/3}. \tag{70}$$

If this formula for t is compared with the corresponding formula (63) for the case of the uniform atmosphere, we observe that the only change is the substitution of k_a for k_0.

The value of M can be found by the method used in the preceding section. We have

$$M = Kk_0(Kk_0a)^{-1/3}\,\frac{\partial}{\partial t}\left[(2\delta)^{1/2}Z\left\{\frac{(2\delta)^{3/2}}{3}\right\}\right]$$

$$= -Kk_0(Kk_0a)^{-1/3}\,\frac{\partial\delta}{\partial t}\left[2\delta + 2\delta Z^2 + \frac{2}{(2\delta)^{1/2}}\,Z\right]$$

$$= -Kk_0(Kk_0a)^{-1/3}\,\frac{\partial\delta}{\partial t}\,[2\delta + \bar{\eta}^2(Kk_0a)^{2/3} + \bar{\eta}(Kk_0a)^{1/3}\delta^{-1}].$$

Now, from (69), $\partial\delta/\partial t = -(k_a/K_{k0})(Kk_0a)^{-1/3}$ so that

$$M = -k_a(Kk_0a)^{-2/3}[2\delta + \bar{\eta}^2(Kk_0a)^{2/3} + \bar{\eta}(Kk_0a)^{1/3}\delta^{-1}]. \tag{71}$$

Again, consider the case where the transmitter and receiver are both on the earth's surface so that $b = r = a$. Using (69) and (71) in (46), we have

$$\begin{aligned} U(a, \theta) &= -\frac{\exp\{ik_a a\theta\}}{\pi a\theta}\frac{[2\pi i(Kk_0a)^{1/3}\theta]^{1/2}}{k_a(Kk_0)^{-1}} \\ &\quad\cdot \sum_i \frac{\exp\{-i(Kk_0a)^{1/3}\theta Kk_0\delta/k_a\}}{2\delta_i + \bar{\eta}^2(Kk_0a)^{2/3} + \bar{\eta}(Kk_0a)^{1/3}\delta^{-1}}e^{i\theta/2} \qquad (72) \\ &= -\frac{\exp\{ik_a d\}}{\pi a\theta}\frac{(2\pi i\chi_1)^{1/2}}{k_a(Kk_0)^{-1}}F_1(\chi_1) \end{aligned}$$

where $\chi_1 = (Kk_0a)^{1/3}\theta$

$$F_1(\chi_1) = \sum_i \frac{\exp\{-i\chi_1 Kk_0\delta/k_a\}}{2\delta_i + \bar{\eta}^2(Kk_0a)^{2/3} + \bar{\eta}(Kk_0a)^{1/3}\delta^{-1}}e^{i\theta/2}. \tag{73}$$

Note that again, if $ak'(a)$ is negligible compared to k_a, so that $K = k_a/k_0$,

$$\chi_1 = (k_a a)^{1/3}\theta, \qquad F_1(\chi_1) = F(\chi)$$

as defined in the preceding section and

$$U(a, \theta) = -\frac{\exp\{ik_a d\}}{\pi a\theta}(2\pi i\chi_1)^{1/2}F(\chi_1). \tag{74}$$

This formula shows that the *Hertz potential in the case of a non-uniform atmosphere can be obtained from the Hertz potential in the case of a constant atmosphere by making the substitutions: k_a for k_0 and $\bar{\eta}$ for η where $\bar{\eta}$ is obtained from η by replacing k_0 by k_a.*

The case where the transmitter and receiver are not on the earth's surface can be derived from the preceding results by multiplying each term in the series for $F(\chi)$ by the corresponding height-gain factors for the transmitter and receiver. The height-gain factor for the receiver is

$$\frac{rw_{t,i}(r)}{aw_{t,i}(a)} = \left(\frac{\tau}{\tau_a}\right)^{1/2}\left[\frac{(ak_a)^2 - \lambda}{(rk)^2 - \lambda}\right]^{1/4}\left(\frac{r}{a}\right)^{1/2}\frac{H_{1/3}^{(1)}(k_0a\tau)}{H_{1/3}^{(1)}(k_0a\tau_a)} \tag{75}$$

by (54). We shall introduce into this formula θ_r, the angular distance of the receiver from the horizon, then (75) will take a simpler form.

From the definition of θ_r, and since $r - a$ is small compared to a, we have $\cos\theta_r = a/r \sim 1 - (r - a)/a$ and since $\cos\theta_r = 1 - \theta_r^2/2$, we have $\theta_r = (2(r - a)/a)^{1/2}$. Put $\chi_r = (ak_a)^{1/3}\theta_r$, then, from (67)

$$\begin{aligned} \left(\frac{k}{k_0}\right)^2 - \frac{\lambda}{(k_0r)^2} \sim 2K^2\left(\frac{r}{r_0} - 1\right) &= 2K^2\left[\frac{r}{a}\{1 + \delta(ak_a)^{-2/3}\}^{-1}\right] \\ &= K^2(ak_a)^{-2/3}(\chi_r^2 + 2\delta) \end{aligned}$$

and we find that $k_0 a\tau = (\chi_r^2 + 2\delta)^{3/2}/3$. Since for $r = a$, $\chi_r = 0$, we see that the right hand member of (75) becomes

$$\left[\frac{\chi_r^2 + 2\delta_i}{2\delta_i}\right]^{1/2} \frac{H_{1/3}^{(1)}\{(\chi_r^2 + 2\delta)^{3/2}/3\}}{H_{1/3}^{(1)}\{(2\delta)^{3/2}/3\}}\,. \tag{76}$$

There is a similar formula for the height gain factor for the transmitter. Instead of χ_r in (76) we must use χ_b where

$$\chi_b = (ak_a)^{1/3}\theta_b\,, \qquad \theta_b = [2(b - a)/a]^{1/2}.$$

It should be pointed out that results similar to those of this section, which show that the field due to a non-uniform atmosphere could be obtained from the field in a uniform atmosphere, have been obtained by Eckersley [10] and Bremmer [11]. However, they had to assume a particular form of variation of the dielectric constant, to wit, $\epsilon(r) = 1 - \alpha + \beta a^2/r^2$ where α and β are small constants. The results of this section are independent of the form and variation f the dielectric constant.

10. *Connection with the "Flat Earth" Theory*

The preceding sections have shown that to a good order of accuracy the eigenvalues depend on the behavior of $k(r)^2$ only in the neighborhood of the transition point, that is, the point at which $k(r)^2 - \lambda/r^2 = 0$. It was also shown that at the transition point $r_0 = a$, very closely. If these facts are taken into account, the differential equation (34) may be approximated as follows:

$$\frac{d^2(rv)}{dr^2} + [k_a^2 N^2 - k_n^2](rv) = 0 \tag{77}$$

where $N^2 = r^2 k(r)^2/a^2 k_a^2$ is the modified index of refraction and where $k_n^2 = t_n(t_n + 1)/a^2$.

Equation (77) is exactly the eigenvalue equation that is obtained in the troposphere theory on the assumption of a flat earth [15]. If the W.K.B. method is used to solve this equation, we see that we get the same results as by applying the W.K.B. method to the more exact equation (34). There will be a discrepancy between the two methods only at points where r differs very much from a. This indicates that the height-gain factors on the "flat earth" theory will be wrong for large heights above the earth's surface. Another possible source of error is due to the fact that the Legendre functions of the exact theory have been approximated by exponentials. This approximation will be bad for extremely large distances between receiver and transmitter. Similar conclusions on the range of validity of the "flat earth" theory were reached by Pekeris [7]. On the whole, however, the results of the previous sections have shown that the "flat earth" theory gives a very satisfactory approximation to the solution.

11. *An Atmosphere Containing a Duct*

In Section 9 we have discussed propagation in a non-uniform atmosphere under the assumption that $r^2k(r)^2$ does not have a stationary point near $r = a$. The reason for this assumption is that, otherwise, the differential equation (34) will have two transition points, that is, the equation $k(r)^2 - \lambda/r^2 = 0$ will have two roots in λ near $\lambda = (ak_a)^2$. Now, the Langer approximation by means of Bessel functions of order one-third is valid in the neighborhood of each transition point separately but it cannot be used in a neighborhood of both transition points. Langer [17] has pointed out that in such a case the approximation to the solutions of the differential equation should be in terms of confluent hypergeometric functions. Since these functions are not too well tabulated, we shall attempt another approach.

In practice, the two transition points of the differential equation are close to each other. We shall approximate the differential equation by assuming that the transition points coincide so that we have a double transition point. In such a case we have $\alpha = 2$ in equation (51) and the approximate solution will be expressed in terms of Bessel functions of order one-fourth.

Let r_0 be a zero of $k(r)^2 - \lambda/r^2 = 0$. Put

$$f(r) = \frac{k(r)^2}{k_0^2a^2} - \frac{\lambda}{k_0^2a^2r^2} - \frac{(r - r_0)}{k_0^2a^2}\left[k(r)^2 - \frac{\lambda}{r^2}\right]'_{r=r_0}$$

and

$$g(r) = (r - r_0)\left[k(r)^2 - \frac{\lambda}{r^2}\right]'_{r=r_0}$$

so that $f(r)$ has a double zero at $r = r_0$. According to (50),

$$\tau = \int_{r_0}^{r} [f(r)]^{1/2}\, dr$$

and then from (52) since $\alpha = 2$, we have that

$$rw_t = \tau^{1/2}[f(r)]^{-1/4}H_{1/4}^{(1)}(k_0a\tau). \tag{78}$$

The eigenvalue equation now becomes

$$af(a)^{1/2}\,\frac{H_{1/4}^{(1)}(k_0a\tau_a)'}{H_{1/4}^{(1)}(k_0a\tau_a)} = -i\,\frac{k_1}{k_0}\left[1 - \left(\frac{k_a}{k_1}\right)^2\right]^{1/2}. \tag{79}$$

In case the earth is assumed to be a perfect conductor, the equation reduces to $H_{1/4}^{(1)}(k_0a\tau_a) = 0$.

The first two roots of the equation $H_{1/4}^{(1)}(\chi) = 0$ are $\chi = (2.38 + 0.17i)\, e^{-\pi i}$, $\chi = (5.6 + 0.17i)\, e^{-\pi i}$, so that the eigenvalue equation becomes

$$k_0a\tau_a = (2.38 + 0.17i)e^{-\pi i}. \tag{80}$$

The integral for τ can be easily approximated in certain cases. If

$$\frac{f'''(r_0)}{f''(r_0)}(r - r_0) \ll 1$$

we may write $f(r) \sim \frac{1}{2} f''(r_0)(r - r_0)^2$. Now

$$f''(r_0) = \left[\frac{(k^2)''}{k_0^2 a^2} - \frac{6\lambda}{k_0^2 a^2 r^4}\right]_{r=r_0} = \frac{2\xi}{a^3},$$

where ξ is defined by the last equation, so that $f(r) = (\xi/a^3)(r - r_0)^2$ and

(81) $$k_0 a \tau_a = k_0 a \int_{r_0}^{a} \frac{\xi^{1/2}}{a^{3/2}}(r - r_0)\, dr = k_0 a(\xi a)^{1/2}\left(\frac{a}{r_0} - 1\right)^2/2.$$

Put, as before in equation (81) $a = r_0(1 + \bar{\delta})$, then $k_0 a \tau_a = \frac{1}{2} k_0 a(\xi a)^{1/2}\bar{\delta}^2$ and in case the earth is a perfect conductor, we must solve the equation:

$$\tfrac{1}{2}(\xi a)^{1/2}\bar{\delta}^2 = (k_0 a)^{-1}(2.38 + 0.17i)e^{-\pi i}.$$

Note that ξ will depend on r_0 and therefore also on $\bar{\delta}$. In case $\xi^{1/2}$ is a slowly varying function of $\bar{\delta}$, we may consider it a constant and then

$$\bar{\delta} = (4.76 + 0.34i)^{1/2} e^{-\pi i/2}(k_0/\xi)^{1/2}(k_0 a)^{-3/4}.$$

Since $t \sim \lambda^{1/2} = k(r_0)r_0 \sim k_0 r_0 \sim k_0 a_0(1 - \delta)$ we find that

(82) $$t \sim k_0 a + (k_0/\xi)^{1/2}(k_0 a)^{1/4}(4.76 + 0.34i)^{1/2} e^{\pi i/2}.$$

Formula (82) shows that the imaginary part of t varies as $(k_0 a)^{1/4}$ for the case of a duct, as contrasted to the case of a homogeneous atmosphere where it varied as $(k_0 a)^{1/3}$. Since the attenuation of the field in decibels depends on the imaginary part of t, we can see that the range will be greatly increased in the case of a duct. For $(k_0 a)$ of the order of 10^8, the range is multiplied by 4.6.

12. *The Electric Dipole*

The preceding work has been carried out for the case of the magnetic dipole because then the equations and the discontinuity conditions become simpler. However, the case of most practical interest is that of an electric dipole. In this section we shall give briefly the equations appropriate for an electric dipole.

First, a new Hertz vector and potential must be defined. Instead of (2) we put

(83) $$\mathbf{H} = \nabla \times \mathbf{C}_1^+$$

where $\mathbf{C}_1^+$ again has only its radial component not zero so that $\mathbf{C}_1 = (rU_1, 0, 0)$. Instead of (6), we find

(84) $$\epsilon \frac{\partial}{\partial r}\frac{1}{\epsilon}\frac{\partial(rU_1)}{\partial r} + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\sin\theta\,\frac{\partial(rU_1)}{\partial\theta} + k^2 rU_1 = 0$$

and instead of (8), we have

$$E_r = \frac{i}{\epsilon\omega}\left(\epsilon\frac{\partial}{\partial r}\frac{1}{\epsilon}\frac{\partial}{\partial r} + k^2\right)(rU_1)$$

$$H_\phi = -\frac{\partial U_1}{\partial\theta}, \qquad E_\theta = -\frac{1}{\epsilon r}\frac{\partial^2}{\partial r\,\partial\theta}(rU_1), \qquad H_r = H_\theta = E_\phi = 0.$$

Since E_θ and H_ϕ must be continuous across a discontinuity surface r = constant, it follows that U_1 and $(1/\epsilon)\,(\partial U_1/\partial r)$ must be continuous across this surface. Instead of (14) we have equation

$$\epsilon\frac{d}{dr}\frac{1}{\epsilon}\frac{d(ru_n)}{dr} + \left[k^2 - \frac{n(n+1)}{r^2}\right](ru_n) = -\frac{\delta(r-b)}{2\pi r} \tag{86}$$

with the same boundary conditions as for (14). Put again $v_n(r)$ for those solutions of the homogeneous equation corresponding to (86), which are regular at the origin, and $w_n^{(1)}(r)$ for those solutions of the homogeneous equation which satisfy the radiation condition at infinity. We find that, now, the Wronskian is no longer a constant, but that

$$W_n(r)/\epsilon_n(r) = \text{constant}. \tag{87}$$

The formula for $U_1(r, \theta)$ is the same as (19):

$$\begin{aligned} U_1(r, \theta) &= \frac{1}{2\pi}\sum_0^\infty \frac{v_n(r)w_n(b)}{W_n(b)}\frac{2n+1}{2}P_n(\cos\theta), \qquad r < b \\ &= \frac{1}{2\pi}\sum_0^\infty \frac{v_n(b)w_n(r)}{W_n(b)}\frac{2n+1}{2}P_n(\cos\theta), \qquad r < b. \end{aligned} \tag{88}$$

The contour integral representation and all the transformations are unchanged. However, the eigenvalue problem becomes the following:

Find those values of t for which there exists a solution of the differential equation,

$$\epsilon\frac{d}{dr}\frac{1}{\epsilon}\frac{d(rv)}{dr} + \left(k^2 - \frac{t(t+1)}{r^2}\right)(rv) = 0 \tag{89}$$

which is both regular at $r = 0$ and satisfies the radiation condition (15) at infinity.

Because of the different discontinuity conditions, the eigenvalue equation is changed. Instead of (41) we have

$$\left.\frac{\dfrac{1}{\epsilon(r)}\dfrac{d}{dr}(rv_t)}{rv_t}\right|_{r=a} = \left.\frac{\dfrac{1}{\epsilon_1}\dfrac{d}{dr}\{r^{1/2}J_{t+\frac{1}{2}}(k_1 r)\}}{r^{1/2}J_{t+\frac{1}{2}}(k_1 r)}\right|_{r=a}$$

so that the eigenvalue equation (44) becomes

$$\left.\frac{(d/dr)(rw_{t_j})}{rw_{t_j}}\right|_{r=a} = -\frac{\epsilon(a)}{\epsilon_1}i(k_1^2 - k^2)^{1/2}. \tag{90}$$

After transforming the contour integral for (88) and using the asymptotic formulas for the Legendre functions, we get the following result, instead of (46)

$$U(r, \theta) = -\sum_j \left[\frac{(2t_j + 1)i}{\pi \sin \theta}\right]^{1/2} \frac{\exp\{i(t_j + \frac{1}{2})\theta\}}{\left.\dfrac{\epsilon(b)}{\epsilon(a)} \dfrac{\partial}{\partial t} \dfrac{(d/dr)(rw_{t_j})}{rw_{t_j}}\right|_a} \frac{w_{t_j}(b)}{aw_{t_j}(a)} \frac{w_{t_j}(r)}{aw_{t_j}(a)} . \tag{91}$$

Here t_j, $j = 1, 2, \cdots$, are the eigenvalues of (89).

Equation (89) can be treated by the W.K.B. method after a transformation is made, to get rid of the ϵ^{-1} factor. Put $rv = \epsilon^{1/2}r\bar{v}$ then equation (89) becomes

$$\frac{d^2}{dr^2}(r\bar{v}) + \left(\bar{k}^2 - \frac{t(t+1)}{r^2}\right)(r\bar{v}) = 0 \tag{92}$$

where $\bar{k}^2 = k^2 + \frac{1}{2}\epsilon^{1/2}(\epsilon^{-3/2}\epsilon')' = k^2 - k\,(d^2/dr^2)(k^{-1})$. This transformation is valid only if we assume ϵ to be continuous and differentiable. Since we shall only apply the W.K.B. method to (89) for $r > a$, our assumption is justifiable.

The eigenvalues t_j of (92) can be found in exactly the same way as we found the eigenvalues for (34). We shall not repeat the work but one difference should be noted. In case the earth is assumed to be a perfect conductor, so $\epsilon_1 = k_1 = \infty$, the eigenvalue equation (90) becomes

$$\left.\frac{(d/dr)(rw_{t_j})}{rw_{t_j}}\right|_{r=a} = 0 \qquad \text{or} \qquad \left.\frac{d}{dr}(rw_{t_j})\right|_{r=a} = 0,$$

instead of $rw_{t_j} = 0$.

Appendix I

The Contour Integral on the Semicircle

In this section we complete the discussion of the contour integral representation in (29) by showing that the integrals over the semi-circle in the first and second quadrant vanish as the radius of the circle goes to infinity. This justifies the expansion of the integral as a sum of residues and an integral over the imaginary axis.

Consider, now, the integral

$$\int \frac{(t + \frac{1}{2})v_t(b)w_t(r)}{W_t} \frac{P_t(-\cos\theta)}{\sin \pi t}\, dt \tag{93}$$

over a quarter circle $|t| = T$ in the first quadrant. Put $t = Te^{i\theta}$, then from (40) we have

$$\left|\frac{P_t(-\cos\theta)}{\sin \pi t}\right| \leq \frac{\exp\{-T\theta \sin\theta\}}{\theta T^{1/2}} .$$

To estimate the other factor in the integrand, we shall investigate the asymptotic behavior for large t of the solutions of the equation corresponding to (16):

$$\frac{d^2(rv)}{dr^2} + \left(k^2 - \frac{t(t+1)}{r^2}\right)(rv) = 0.$$

It can be easily verified by substitution in the differential equation that there are two independent solutions which for large t behave either like

$$rv \sim C_1 r^{t+1}\left[1 - \frac{1}{2t+1}\int_0^r k^2 r\, dr\right] \qquad \text{or} \qquad rv \sim C_2 r^{-t}\left[1 + \frac{1}{2t}\int_0^r k^2 r\, dr\right]$$

where C_1 and C_2 are independent of r. Since rv_t was by definition regular at $r = 0$, it must behave like $C_1 r^{t+1}$ for large t. When these asymptotic formulas are differentiated with respect to t, we find that for all solutions of the differential equations

$$\left|\frac{(d/dr)(rv)}{rv}\right| < \frac{t}{r}. \tag{94}$$

Now

$$\frac{(t+\frac{1}{2})v_t(b)w_t(r)}{W_t} = \frac{(t+\frac{1}{2})bv_t(b)rw_t(r)}{br\left[rw_t(r)\dfrac{d}{dr}(rv_t(r)) - rv_t(r)\dfrac{d}{dr}(rw_t(r))\right]}$$

$$= \frac{t+\frac{1}{2}}{br}\frac{[bv_t(b)]/[rv_t(r)]}{M} \qquad \text{where} \qquad M = \frac{(d/dt)(rv_t)}{rv_t} - \frac{(d/dt)(rw_t)}{rw_t}.$$

M will vanish if and only if $W_t = 0$. If T is so chosen that the quarter circle does not go through any zero of W_t, then from (94) M will be bounded below by some multiple of t/r. We have then

$$\frac{(t+\frac{1}{2})v_t(b)w_t(r)}{W_t} \sim \frac{Ct}{br}\left(\frac{b}{r}\right)^{t+1}\frac{1}{t/r} = \frac{C}{b}\left(\frac{b}{r}\right)^{t+1}$$

and since $r < b$, this term will go to zero as T goes to infinity. Finally, the integral will be less in absolute value than

$$\frac{C}{b\theta T^{1/2}}\int_0^{\pi/2} \exp\{-T\theta\sin\theta\}\, Td\theta \left(\frac{b}{r}\right)^{T\cos\theta}$$

and as this clearly approaches zero for large T, the shifting of the contour in section 4 was justified.

Appendix II

An Estimate for the Integral

It has been shown that the contour integral representation of the Hertz potential can be expressed (39) as a sum of residues plus an integral over the imaginary axis. In this section we estimate the value of the integral and show two things:

First, the integral vanishes identically if the earth is assumed to be a perfect conductor; second, the integral is small compared to the first term in the residue series if the complex dielectric constant of the earth is small compared to the dielectric constant of the atmosphere. Since this condition is always satisfied in practice, we are justified in neglecting the contribution of the integral.

In order to estimate the integrand in (39) we must make use of the fact that the differential equation (34) is unchanged when t is replaced by $-t-1$ so that

$$w_t^{(1)}(r) = w_{-t-1}^{(1)}(r) \qquad \text{and} \qquad w_t^{(2)}(r) = w_{-t-1}^{(2)}(r).$$

Since $w_t^{(1)}$ and $w_t^{(2)}$ are independent solutions of (34), we may write

$$v_t(r) = \alpha w_t^{(1)}(r) + \beta w_t^{(2)}(r)$$

where α and β are independent of r but may depend on t.

Note that $w_t = \beta$ Wronskian $(w_t^{(2)}, w_t^{(1)}) = 2ik_0\beta$. Now, we make use of the fact that for $r < a$, $k^2 = k_1^2$, is a constant so that

$$rv_t(r) = r^{1/2}J_{t+\frac{1}{2}}(k_1 r), \qquad r < a.$$

At the boundary $r = a$, we must have

$$\alpha a w_t^{(1)}(a) + \beta a w_t^{(2)}(a) = a^{1/2}T_{t+\frac{1}{2}}(k_1 a)$$

$$\alpha \frac{d}{dr}(rw_t^{(1)})\bigg|_a + \beta \frac{d}{dr}(rw_t^{(2)})\bigg|_a = \frac{d}{dr}(r^{1/2}J_{t+\frac{1}{2}})\bigg|_a .$$

These two equations will determine α and β. We find

$$\frac{\alpha}{\beta} = -\left.\frac{rw_t^{(2)}\dfrac{d}{dr}(r^{1/2}J_{t+\frac{1}{2}}) - r^{1/2}J_{t+\frac{1}{2}}\dfrac{d}{dr}(rw_t^{(2)})}{rw_t^{(1)}\dfrac{d}{dr}(r^{1/2}J_{t+\frac{1}{2}}) - r^{1/2}J_{t+\frac{1}{2}}\dfrac{d}{dr}(rw_t^{(1)})}\right|_{r=a}$$

$$= -\left.\frac{rw_t^{(2)}(r)\dfrac{d}{dr}[\log(r^{1/2}J_{t+\frac{1}{2}})] - \dfrac{d}{dr}[\log(rw_t^{(2)})]}{rw_t^{(1)}(r)\dfrac{d}{dr}[\log(r^{1/2}J_{t+\frac{1}{2}})] - \dfrac{d}{dr}[\log(rw_t^{(1)})]}\right|_{r=a} \tag{95}$$

$$= -\frac{aw_t^{(2)}(a)}{aw_t^{(1)}(a)}M_t ,$$

where M_t is defined by the last equation. Now, consider the expression

$$\frac{v_t(b)w_t(r)}{W_t} = \frac{[\alpha w_t^{(1)}(b) + \beta w_t^{(2)}(b)]w_t^{(1)}/(r)}{2ik_0\beta}$$

$$= \frac{w_t^{(2)}(b)w_t^{(1)}(r)}{2ik_0} - \frac{w_t^{(1)}(a)}{2ik_0 w_t^{(1)}(a)} w_t^{(1)}(b)w_t^{(1)}(r)M_t \; .$$

Everything in this formula, except the factor M_t , is unchanged when t is replaced by $-t - 1$, so that

$$\frac{v_t(b)w_t(r)}{W_t} - \frac{v_{-t-1}(b)w_{-t-1}(r)}{W_{-t-1}} = -\frac{w_t^{(2)}(a)w_t^{(1)}(b)w_t^{(1)}(r)}{2ik_0 w_t^{(1)}(a)} [M_t - M_{-t-1}].$$

Now

$$M_t = \frac{\left[\frac{d}{dr}\log(r^{1/2}J_{t+\frac{1}{2}})\right]_a - \left[\frac{d}{dr}\log rw_t^{(2)}\right]_a}{\left[\frac{d}{dr}\log(r^{1/2}J_{t+\frac{1}{2}})\right]_a - \left[\frac{d}{dr}\log rw_t^{(1)}\right]_a}$$

(96)

$$= \frac{\left[\frac{d}{dr}\log rw_t^{(1)}\right]_a - \left[\frac{d}{dr}\log rw_t^{(2)}\right]_a}{\left[\frac{d}{dr}\log(r^{1/2}J_{t+\frac{1}{2}})\right]_a - \left[\frac{d}{dr}\log rw_t^{(1)}\right]_a} + 1$$

and we see that in this formula the only term which changes when t is replaced by $-t - 1$ is

$$\frac{d}{dr}[\log(r^{1/2}J_{t+\frac{1}{2}})]_a \; .$$

However, the argument of this Bessel function is k_1a, which is a very large quantity since the conductivity of the earth is so large, and by the asymptotic formula for the Bessel functions of large argument we have

$$\left[\frac{d}{dr}\log(r^{1/2}J_{t+\frac{1}{2}})\right]_a = k_1\tan\left(k_1a - t\frac{\pi}{2} - \frac{\pi}{2}\right).$$

Note that this formula is valid only if k_1a is very much larger than t. Now since k_1 is larger than k, the fraction in the formula (96) must be very small so that $M_t - M_{-t-1}$ is nearly zero.

If we put t equal to $i\tau - \frac{1}{2}$ so that $-t - 1 = -i\tau - \frac{1}{2}$, we see that the bracket in the integrand of (39) must be nearly zero for τ very much smaller than k_1a. In the remaining interval of integration we use the formula (40). We find then that the integrand is of the order of magnitude $e^{-\tau\theta}$. Now

$$\int_{ka}^{\infty} e^{-\tau\theta}\,d\tau = \frac{1}{\theta}e^{-ka\theta}$$

and we see that this is very small compared to exp $\{-(ka)^{1/3}\theta\}$ or to exp $\{-(ka)^{1/4}a\}$ which is what is obtained from the residue series. We have thus shown that the integral in (39) is small compared to the first term in the residue series.

Bibliography

1. Watson, G. N., *The diffraction of electrical waves by the earth,* Proceedings of the Royal Society of London, Volume A95, 1918–1919, p. 83.
2. Bremmer, H., and van Der Pol, B., *The diffraction of electro-magnetic waves from an electrical point source round a finitely conducting sphere, with applications to radiotelegraphy and the theory of the rainbow,* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Volume 24, 1937, p. ?41.
3. Rydbeck, Olof E. H., *On the propagation of radio waves,* Transactions Chalmers University of Technology, No. 34, 1944.
4. Pryce, M. H., unpublished work.
5. Booker, H. G., and Walkinshaw, W., *The mode theory of tropospheric refraction and its relation to wave-guides and diffraction,* Meteorological Factors in Radio Wave Propagation. Report of a conference held on April 8, 1946, by the Physical Society and The Royal Meteorological Society. The Physical Society, London, 1946, p. 80.
6. Furry, W. H., *Theory of characteristic functions in problems of anomalous propagation,* O.E.M. Sr-262, Division 14, Report 680, Radiation Laboratory, M.I.T., Feb. 28, 1945.
7. Pekeris, C. L., *Accuracy of the earth flattening approximation in the theory of microwave propagation,* Report M.P.G.-3 Columbia University Mathematical Physics Group, April 15, 1946.
8. Schelling, I. C., Burrows, C. R., and Ferrell, E. B., *Ultra short wave propagation,* Proceedings of the Institute of Radio Engineers, Volume 21, 1933, p. 427.
9. Langer, R. E., *On the asymptotic solutions of ordinary differential equations, with an application to the Bessel functions of large order,* Transactions of the American Mathematical Society, Volume 33, 1931, p. 23.
10. Eckersley, T. L., and Millington, G., *Application of the phase integral method to the analysis of the diffraction and refraction of wireless waves round the earth,* Philosophical Transactions of the Royal Society of London, Volume A237, 1938, p. 273.
11. Bremmer, H., *Terrestrial Radio Waves.* Elsevier Publishing Company, New York–Houston, 1949.
12. Stratton, J. A., *Electromagnetic Theory,* McGraw-Hill, New York, 1941, p. 415. Frank, P., and von Mises, R., *Die Differential und Integralgleichungen der Mechanik und Physik,* Vol. II, Vieweg, Braunschweig, 1930, p. 871.
13. Stratton, J. A., *Electromagnetic Theory,* McGraw-Hill, New York, 1941, p. 431.
14. Hobson, E. W., *The Theory of Spherical and Ellipsoidal Harmonics,* University Press, Cambridge, 1931.
15. Pekeris, C. L., *Wave theoretical interpretation of propagation of 10-centimeter and 3-centimeter waves in low-level ocean ducts,* Proceedings of the Institute of Radio Engineers, Volume 35, 1947, p. 453.
16. Langer, R. E., *The asymptotic solutions of certain linear ordinary differential equations of the second order,* Transactions of the American Mathematical Society, Volume 36, 1934, p. 90.

Reflection of Electromagnetic Waves from Slightly Rough Surfaces

By STEPHEN O. RICE
Bell Telephone Laboratories

I. *Introduction*

The problem of reflection of waves from a rough or corrugated surface is of interest in a number of different fields. In particular, the problem occurs in the propagation of radio waves over rough ground or over the sea. The general problem of reflection from rough surfaces appears to be difficult. However, a number of investigators have been able to make progress by dealing with special cases or by making suitable assumptions. For example, the radio problem in which the impressed field originates at a point has been studied by E. Feinberg [1]. L. V. Blake [2] has applied probability theory to the problem of calculating the reflection of radio waves from a rough sea. The reflection from the very rough surface formed by the edges of an infinite set of plates has been investigated by J. F. Carlson and A. E. Heins [3]. Also, in addition to the studies of W. S. Ament[1] mentioned below, I understand that a considerable amount of work on this subject, which is unpublished as yet, has been done by Mr. Twersky [7] of New York University, and by Messrs. Norton and Hufford and their associates at the National Bureau of Standards.

Here we shall be concerned with the reflection of plane electromagnetic waves from a surface $z = f(x, y)$ which is almost, but not quite, flat. The small deviations of this surface from the x,y-plane are of a random nature. Except in Section 7, the surface is assumed to be a perfect conductor. Although in practical cases the surfaces are usually much rougher than is assumed here, our problem has the virtue of being one of the simplest which still shows the effect of roughness.

The roughness of the surface is described by a "roughness spectrum" or "roughness distribution function" $W(p, q)$. When the surface is expressed as

Paper presented at the June, 1950, Symposium on the Theory of Electromagnetic Waves, under the sponsorship of the Washington Square College of Arts and Science and the Institute for Mathematics and Mechanics of New York University and the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories.

[1]I am indebted to Mr. Ament of the Naval Research Laboratory for an opportunity to study some of his work before its publication. I also wish to acknowledge the help I have received from discussions of the general problem of reflection with K. Bullington of the Bell Telephone Laboratories.

the sum of two-dimensional Fourier components, $W(p, q)\,dp\,dq$ represents the relative strength (as measured by their contribution to the mean square value of $f(x, y)$) of those components which go through between p and $p + dp$ radians/meter in the x direction and through between q and $q + dq$ in the y direction. If the average distance between the hills on the surface is large, and the surface is smooth except for these hills, $W(p, q)$ will be appreciably different from zero only for small values of p and q. The associated auto-correlation function of the surface, which may be expressed as the Fourier transform of $W(p, q)$, is not used here although it does occur in the work of W. S. Ament.

The reflected field is determined by a method similar to that used by Rayleigh [4] to study the reflection of acoustic waves from rough walls. The expressions which we obtain for the field are not exact since the boundary conditions at the surface are satisfied only to within $O(f^2(x, y))$, i.e. to within terms of the second order—a shortcoming forced upon us by the increasing complexity of our successive approximations. The two cases corresponding to horizontal polarization (incident E vector parallel to x,y-plane) and vertical polarization (incident H vector parallel to x,y-plane), respectively, are considered.

After expressions for the components of the field are obtained, various averages are computed, the average being taken over many surfaces which are different but which have the same statistical properties. In particular, the average value of the reflected field leads to an expression for the reflection coefficient. It is found that this reflection coefficient depends upon the polarization in somewhat the same way as does the reflection coefficient for an almost, but not quite, perfectly conducting plane. Also, when the average distance between hills is large, the reflection coefficients for both the horizontal and vertical polarizations reduce to the same expression. By a method similar to the one used in the study of Fraunhofer diffractions, W. S. Ament has obtained an expression for the average reflection coefficient when the distance between hills is large and they are such that they do not cast any shadows (with respect to the incident wave). Our approximate expression agrees with the first two terms in the expansion of Ament's expression, which is as much of an agreement as the accuracy of our work allows.

Closely associated with the problem of reflection is the problem of surface wave propagation. This corresponds loosely to the case of grazing incidence and vertical polarization; a modified form of the reflection analysis may be used to obtain an expression for the propagation constant of the surface wave. It is found that, roughly speaking, the Fourier components of the surface whose wavelengths are much greater than that of the electromagnetic wave tend to produce attenuation through scattering, while the guiding action of the surface is due to the components of shorter wavelength. This is in accord with the results of earlier studies of surface waves on corrugated surfaces [5, 6].

The method used to study reflection from a slightly rough but perfectly conducting surface may be extended to take into account the electrical properties of the reflecting medium. This is done in Section 7 for the case of hori-

zontal polarization, the magnetic permeabilities of the two media being assumed equal. There are two reasons for this study. The first is to determine the additional amount of complication introduced. The second is to show that an annoying difficulty encountered in the perfect conductor case, namely that the integral for the mean square value of E_z (i.e. the component of electric intensity which is approximately normal to the surface) sometimes diverges logarithmically, may be removed by taking into account the finite conductivity of the reflector.

2. *Description of Rough Surface*

We shall take the equation of the perfectly conducting rough surface to be

$$z = f(x, y) = \sum_{mn} P(m, n) \exp \{-ia(mx + ny)\}$$

(2.1)

$$a = 2\pi/L$$

where the double summation extends from $-\infty$ to $+\infty$ for both m and n. The definition of a shows that $f(x, y)$ is periodic in both x and y with period L (assumed to be large). In order to make $f(x, y)$ real we impose the condition

(2.2) $$P(-m, -n) = P^*(m, n)$$

where the asterisk denotes the conjugate complex quantity.

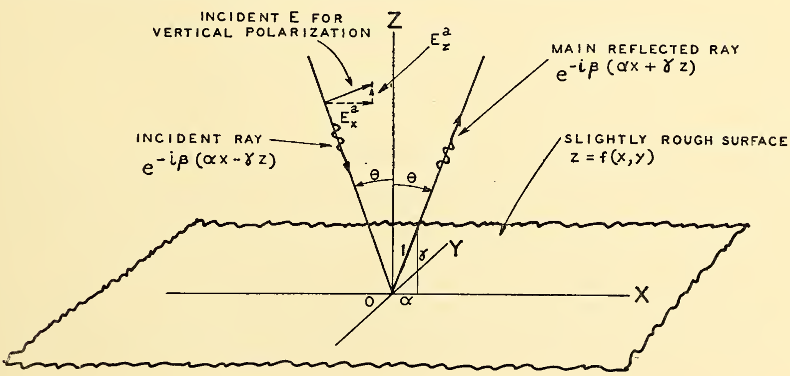


FIG. 1. Diagram showing the coordinate system and the incident E vector for vertical polarization. For horizontal polarization the incident E vector is parallel to the y axis.

The random character of roughness is introduced by taking the coefficients $P(m, n)$ to be independent random variables, subject only to (2.2). For the sake of being definite, we assume $P(0, 0)$ and the real and imaginary parts of $P(0, 1)$, $P(1, 0)$, $P(2, 0)$, $P(1, 1)$, $P(0, 2)$, $P(1, -1)$, etc. to be independent random variables distributed normally about zero. We assume further that, for assigned values of m and n, the four independent random variables formed by the real and imaginary parts of $P(m, n)$ and $P(m, -n)$ all have the same vari-

ance, i.e., the same mean square value. When we use angular brackets to denote average values, our assumptions tell us that

$$\langle P(m, n)\rangle = 0$$

$$\langle P(m, n)P(u, v)\rangle = 0, \qquad (u, v) \neq (-m, -n)$$

$$\langle P(m, n)P^*(m, n)\rangle = \langle P(m, n)P(-m, -n)\rangle = \pi^2 W(p, q)/L^2 \tag{2.3}$$

$$W(p, q) = W(|p|, |q|)$$

$$p = am = 2\pi m/L, \qquad q = an = 2\pi n/L$$

Here $\langle\ \rangle$ denotes that m and n are to be held fixed and the average taken over the universes of the real and imaginary parts of the $P(m, n)$'s. $W(p, q)$ is the roughness spectrum mentioned in the introduction, and p and q are radian wave numbers. Note that $\langle P^2(m, n)\rangle$ is zero, except when $m = n = 0$, by virtue of the real and imaginary parts of $P(m, n)$ having the same variance. Incidentally, the statistical properties of $P(m, n)$ were obtained by expressing the typical Fourier series term

$$\begin{aligned}&(a_{mn} \cos amx + b_{mn} \sin amx) \cos any \\ &\qquad + (c_{mn} \cos amx + d_{mn} \sin amx) \sin any, \qquad m > 0, n > 0\end{aligned} \tag{2.4}$$

as the sum of four exponential terms. This leads to four relations of the form

$$P(m, n) = \frac{a_{mn} + ib_{mn} + ic_{mn} - d_{mn}}{4}$$

and the properties of $P(m, n)$ follow when $a_{mn}, \cdots, d_{mn}$ are assumed to be independent random variables distributed normally about zero with the same variance, namely $4\pi^2 W(p, q)/L^2$. The $4\pi^2$ arises from the fact that we have elected to measure p and q in radians/meter instead of cycles/meter.

Equation (2.1) defines a surface for each set of coefficients. As an example of the use we shall make of $W(p, q)$ we compute the average value of $f^2(x, y)$ as we go from surface to surface, holding x and y fixed all the while.

$$\begin{aligned}\langle f^2(x, y)\rangle &= \sum_{mnuv} \langle P(m, n)P(u, v)\rangle \exp\{-iax(m + u) - iay(n + v)\} \\ &= \sum_{mn} \langle P(m, n)P(-m, -n)\rangle \\ &\to \int_{-\infty}^{\infty} dm \int_{-\infty}^{\infty} dn\, \pi^2 W(p, q)/L^2 \\ &= \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, \frac{W(p, q)}{4} = \int_0^{\infty} dp \int_0^{\infty} dq\, W(p, q).\end{aligned} \tag{2.5}$$

Here we have used (2.3). In going from the summation to the double integral we have let the period L approach infinity. It is seen that $W(p, q)\, dp\, dq/4$ represents the contribution to $\langle f^2(x, y)\rangle$ of those components in (2.1) having between p and $p + dp$ radians/meter in the x direction and between q and $q + dq$ radians/meter in the y direction.

3. *Incident Wave Horizontally Polarized*

We assume the components of the total electric intensity for $z > f(x, y)$ to be

$$
\begin{aligned}
E_x &= \sum A_{mn} E(m, n, z) \\
E_y &= 2i \sin \beta\gamma z \exp\{-ia\nu x\} + \sum B_{mn} E(m, n, z) \qquad (3.1) \\
E_z &= \sum C_{mn} E(m, n, z)
\end{aligned}
$$

where the summations extend from $-\infty$ to $+\infty$ for m and n and

$$
E(m, n, z) = \exp\{-ia(mx + ny) - ib(m, n)z\}. \tag{3.2}
$$

The time factor $\exp\{i\omega t\}$ is understood. $b(m, n)$ is either a positive real or a negative imaginary number:

$$
b(m, n) = \begin{cases} [\beta^2 - a^2m^2 - a^2n^2]^{1/2}, & m^2 + n^2 < \beta^2/a^2 \\ -i[a^2m^2 + a^2n^2 - \beta^2]^{1/2}, & m^2 + n^2 > \beta^2/a^2 \end{cases} \tag{3.3}
$$

where $\beta = 2\pi/\lambda$, λ being the wavelength of the incident wave. A_{mn}, B_{mn}, C_{mn} are constants which we shall determine approximately on the assumption that βf and the partial derivatives f_x and f_y are small (here and in what follows we shall often denote $f(x, y)$ by f) compared to unity.

The field obtained from (3.1) when the summations are omitted is the one which would occur if the perfectly conducting surface were flat ($f \equiv 0$). In (3.1) we take ν to be an integer so that the field is periodic in x and y of period L by virtue of $a = 2\pi/L$. It follows that the angle θ between the incoming ray and the z-axis is restricted to certain discrete values given by

$$
\begin{aligned}
a\nu &= 2\pi\nu/L = \beta \sin\theta = \beta\alpha \qquad |a\nu| < \beta \\
\alpha &= \sin\theta, \qquad \gamma = \cos\theta \qquad 0 \le \gamma.
\end{aligned} \tag{3.4}
$$

Since L becomes very large we can pick an integer ν which will correspond approximately to any angle of incidence. The leading term in E_y may be written as

$$
\exp\{-i\beta(\alpha x - \gamma z)\} - \exp\{-i\beta(\alpha x + \gamma z)\}
$$

where the first term represents the incoming wave and the second term the

main part of the reflected wave. It is seen that the direction cosines of the incident and reflected rays are $(\alpha, 0, -\gamma)$ and $(\alpha, 0, \gamma)$, respectively. From the definition of $b(m, n)$ it follows that

$$b(\nu, 0) = (\beta^2 - a^2\nu^2)^{1/2} = \beta\gamma. \tag{3.5}$$

The exponential form (3.2) of $E(m, n, z)$ ensures that all three components of the electric intensity (3.1) satisfy the wave equation. The coefficients are determined by the relation div $E = 0$, which gives

$$amA_{mn} + anB_{mn} + b(m, n)C_{mn} = 0, \tag{3.6}$$

together with the condition that the tangential component of E must vanish at the perfectly conducting surface $z = f$. If N denotes the unit vector normal to the surface, $N(E \cdot N)$ is the component of the electric intensity normal to the surface. The remaining portion of E, the tangential component, is $E - N(E \cdot N)$, all three components of which must vanish. Equating the x and y components to zero gives

$$\begin{aligned} E_x - N_x(E_xN_x + E_yN_y + E_zN_z) &= 0 \\ E_y - N_y(E_xN_x + E_yN_y + E_zN_z) &= 0 \end{aligned} \tag{3.7}$$

If these two equations are satisfied the z component is also zero (if $N_z \neq 0$) as may be seen by multiplying the first by N_x, the second by N_y, and adding. The components of N are

$$N_x = -f_xN_z, \qquad N_y = -f_yN_z, \qquad N_z = (1 + f_x^2 + f_y^2)^{-1/2}. \tag{3.8}$$

We now assume βf, f_x, f_y all to be of the same order of smallness which, for the sake of simplicity, we shall denote by $O(f)$ instead of $O(\beta f)$. Likewise instead of $O(\beta^2 f^2)$ we shall write $O(f^2)$, and so on. In our work we shall neglect $O(f^3)$ terms and it will not be necessary to go beyond the leading terms in

$$N_x = -f_x + O(f^3), \qquad N_y = -f_y + O(f^3), \qquad N_z = 1 + O(f^2) \tag{3.9}$$

Near the surface $z = f$, i.e. near $z = 0$, the leading term in E_y as given by (3.1) is $O(f)$, and we assume for the moment that E_x and E_z are also of this order. Then, neglecting $O(f^3)$ terms in (3.7), we obtain the two boundary conditions

$$\begin{aligned} E_x - N_xE_z &= 0 \\ E_y - N_yE_z &= 0 \end{aligned} \tag{3.10}$$

which must hold at $z = f$. Thus, if E_z is $O(f)$, then both E_x and E_y must be $O(f^2)$ at the surface. Of course this holds only for horizontal polarization. For vertical polarization it turns out that E_z is $O(1)$ and both E_x and E_y are $O(f)$.

The problem now is to choose the coefficients in (3.1) so that the divergence relation (3.6) and the boundary conditions (3.10) are satisfied to within $O(f^2)$.

Writing

$$\sin \beta\gamma f = \beta\gamma f + O(f^3)$$

$$(3.11) \qquad E(m, n, f) = [1 - ib(m, n)f + \cdots]E(m, n, 0)$$

$$A_{mn} = A_{mn}^{(1)} + A_{mn}^{(2)} + \cdots$$

where $A_{mn}^{(1)}$ is $O(f)$, $A_{mn}^{(2)}$ is $O(f^2)$, etc., and expressing B_{mn}, C_{mn} in a similar way enables us to write the boundary conditions (3.10) as

$$\sum [A_{mn}^{(1)} + A_{mn}^{(2)} + f_x C_{mn}^{(1)}][1 - ib(m, n)f]E(m, n, 0) = 0$$

$$(3.12) \qquad 2i \exp \{-iavx\} \cdot \beta\gamma f$$

$$+ \sum [B_{mn}^{(1)} + B_{mn}^{(2)} + f_y C_{mn}^{(1)}][1 - ib(m, n)f]E(m, n, 0) = 0$$

where we have neglected $O(f^3)$ terms. In this work we shall overlook questions of convergence although they may perhaps be treated by placing suitable restrictions on the components $P(m, n)$ of the surface f.

Equating the first order terms in (3.12) to zero gives

$$(3.13) \qquad \begin{aligned} &\sum A_{mn}^{(1)} E(m, n, 0) = 0 \\ &2i \exp \{-iavx\}\beta\gamma f + \sum B_{mn}^{(1)} E(m, n, 0) = 0. \end{aligned}$$

Likewise, the second order terms yield

$$(3.14) \qquad \begin{aligned} &\sum [A_{mn}^{(2)} + f_x C_{mn}^{(1)} - ib(m, n)fA_{mn}^{(1)}]E(m, n, 0) = 0 \\ &\sum [B_{mn}^{(2)} + f_y C_{mn}^{(1)} - ib(m, n)fB_{mn}^{(1)}]E(m, n, 0) = 0. \end{aligned}$$

As (3.2) shows, $E(m, n, 0)$ is the exponential function of x and y which occurs in a double Fourier series. Hence the first of equations (3.13) requires $A_{mn}^{(1)} = 0$. In order to interpret the remaining equations we need the following results. Writing u, v for m, n in (2.1) and using the definition of $E(m, n, z)$ leads to

$$(3.15) \qquad \begin{bmatrix} f \\ f_x \\ f_y \end{bmatrix} = \sum_{uv} \begin{bmatrix} 1 \\ -iau \\ -iav \end{bmatrix} P(u, v)E(u, v, 0)$$

whence, upon setting $m = u + \nu$, $n = v$,

$$(3.16) \qquad \begin{aligned} \exp \{-iavx\} f &= \sum_{uv} P(u, v)E(u + \nu, v, 0) \\ &= \sum_{mn} P(m - \nu, n)E(m, n, 0). \end{aligned}$$

A somewhat similar argument may be used to establish

$$\sum_{mn} \begin{bmatrix} f \\ f_x \\ f_y \end{bmatrix} J_{mn} E(m, n, 0)$$

(3.17)

$$= \sum \begin{bmatrix} 1 \\ -ia(m-k) \\ -ia(n-l) \end{bmatrix} J_{kl} P(m-k, n-l) E(m, n, 0)$$

where the summation for m, n, k, l on the right extends from $-\infty$ to ∞ and J_{mn} represents an arbitrary function of m and n. (3.17) is obtained by replacing m, n by k, l on the left and then introducing (3.15). The two E functions may be combined by the multiplication law for the exponential function and the right side of (3.17) obtained upon setting $m = u + k$, $n = v + l$.

Equating the coefficient of $E(m, n, 0)$ to zero in the second of equations (3.13) after using (3.16) gives

$$B_{mn}^{(1)} = -2i\beta\gamma P(m - \nu, n). \tag{3.18}$$

The second order terms $A_{mn}^{(2)}$, $B_{mn}^{(2)}$ may be obtained by setting the values of $A_{mn}^{(1)}$, $B_{mn}^{(1)}$ in (3.14) and using (3.17):

$$\begin{aligned} A_{mn}^{(2)} &= \sum_{kl} ia(m-k) C_{kl}^{(1)} P(m-k, n-l) \\ B_{mn}^{(2)} &= \sum_{kl} [ia(n-l) C_{kl}^{(1)} + 2\beta\gamma b(k, l) P(k-\nu, l)] P(m-k, n-l). \end{aligned} \tag{3.19}$$

Once the A's and B's are known the C's may be obtained from the divergence relation (3.6). For example

$$C_{mn}^{(1)} = -anB_{mn}^{(1)}/b(m, n) = 2i\beta\gamma anP(m-\nu, n)/b(m, n). \tag{3.20}$$

When the appropriate expressions for the coefficients are set in (3.1) we get

$$\begin{aligned} E_x &= -2\beta\gamma \sum_{mn} E(m, n, z) \sum_{kl} a^2(m-k) l Q(m, n, k, l) \\ E_y &= 2i \exp\{-i\beta\alpha x\} \sin \beta\gamma z - 2\beta\gamma \sum_{mn} E(m, n, z)[iP(m-\nu, n) \\ &\quad + \sum_{kl} \{a^2(n-l)l - b^2(k, l)\} Q(m, n, k, l)] \\ E_z &= 2\beta\gamma \sum_{mn} [E(m, n, z)/b(m, n)][ianP(m-\nu, n) \\ &\quad + \sum_{kl} \{a^3 l(m^2 + n^2 - mk - nl) - anb^2(k, l)\} Q(m, n, k, l)] \end{aligned} \tag{3.21}$$

where $E(m, n, z)$ is the exponential function defined by (3.2), $(\alpha, 0, -\gamma)$ are direction cosines of the incident ray, $a = 2\pi/L$ where L is the period of the surface, ν is an integer given by $a\nu = \beta\alpha$, $\beta = 2\pi/\lambda$ and

$$(3.22) \qquad Q(m, n, k, l) = P(k - \nu, l)P(m - k, n - l)/b(k, l)$$

The summations for m, n, k, l extend from $-\infty$ to $+\infty$.

The terms entering the summations in (3.21) may be divided into two classes. A term is in the first class if the corresponding values of m and n satisfy $a^2m^2 + a^2n^2 < \beta^2$ and in the second if the opposite inequality is satisfied. For a term in the first class $b(m, n)$ is positive real and $E(m, n, z)$ represents a wave traveling in the direction specified by the direction cosines am/β, an/β, $b(m, n)/\beta$. The corresponding electric intensity is perpendicular to the direction of propagation, as is shown by (3.6). In terms of the wavelength λ of the incident wave and the period L of the surface these direction cosines are $\lambda m/L$, $\lambda n/L$, $b(m, n)/\beta$. For a term in the second class $b(m, n)$ is negative imaginary and $E(m, n, z)$ corresponds to a surface wave traveling in the direction determined by m/n and exponentially attenuated in the z direction.

An examination of the series for E_z shows that the terms become large for n near $\pm\beta/a = \pm L/\lambda$ and for m near zero if the coefficients around $P(-\nu, \beta/a)$ are appreciably different from zero, for then $b(m, n)$ in the denominator is small. This indicates that for some surfaces there will be an appreciable sidewise (i.e., in the y direction) scattering of the wave. It will be seen later that if the finite conductivity of the reflecting surface is taken into account the large terms remain finite even if $b(m, n) = 0$.

That E_z sometimes tends to be large may be seen from the following physical considerations. Take the case of normal incidence so that $\nu = 0$ and take the surface to be $z = 2P \cos \beta y$. The incident E_y produces a surface current in the y direction and each upward (and downward) slope of the surface may be regarded as a surface current element (infinitely long in the x direction) which radiates a field. Since the period of the surface is equal to one wavelength at the incident radiation, the E_z components of the fields of various current elements are in phase at $z = 0$ and hence the resultant E_z tends to be large.

As the roughness increases, the additional energy in the scattered radiation is obtained at the expense of the energy in the main component of the reflected wave. This is closely connected with the relation

$$(3.23) \qquad \text{Real Part of } 2\beta\gamma B_{\nu 0} = \sum [| A^2_{mn} | + | B^2_{mn} | + | C^2_{mn} |]b(m, n)$$

which is an extension of a result due to Rayleigh [4]. Here the summation extends over all values of the integers m and n such that $m^2 + n^2 < \beta^2/a^2$ (i.e., over the values for which $b(m, n)$ is real). Equation (3.23) is an exact relation and does not depend on $z = f(z, y)$ being only slightly rough. It may be established by equating to zero the average power flow through a square of side L lying on a plane $z =$ constant parallel to the x,y-plane and at a great height above it. $B_{\nu 0}$ is the change in the main reflected wave produced by the rough-

ness. Although (3.23) was derived by integrating Poynting's vector over the square, it is interesting to note that the m,n-th term on the right is proportional to the intensity of the m,n-th component of the field times the cosine, $b(m, n)/\beta$, of the angle between its direction of propagation and the z-axis. That (3.23) and (3.21) are in accord may be seen from the fact that (since $iP(\nu - \nu, 0)$ is imaginary) the real part of $B_{\nu 0}$ is, from (3.21) and (3.1),

$$(3.24) \qquad 2\beta\gamma \sum [a^2l^2 + b^2(k, l)] \mid P(k - \nu, l) \mid^2/b(k, l) + O(f^3)$$

where the summation extends over values of k and l such that $k^2 + l^2 < \beta^2/a^2$ (because $b(k, l)$ is real for only these values). Furthermore,

$$\mid A_{mn} \mid^2 = O(f^4), \qquad \mid B_{mn} \mid^2 = \mid 2\beta\gamma P(m - \nu, n) \mid^2 + O(f^3)$$

$$\mid C_{mn} \mid^2 = \mid 2\beta\gamma a n P(m - \nu, n)/b(m, n) \mid^2 + O(f^3)$$

and when these are put in the right hand side of (3.23) we get a result which agrees with (3.24).

Up to this point the results of this section hold for any assigned values of the $P(m, n)$'s except that they are usually required to be small. No statistical considerations enter into equations (3.1) to (3.21). However, from here to the end of this section we shall make use of the statistical properties of the $P(m, n)$'s, described in Section 2, to obtain various average values from the approximate expressions (3.21) for the field. From (2.3) and (3.22) follows

$$(3.25) \qquad \langle Q(m, n, k, l)\rangle = \begin{cases} 0, & (m, n) \neq (\nu, 0) \\ \pi^2 W(ak - a\nu, al)/L^2 b(k, l), & m = \nu, n = 0. \end{cases}$$

When the averages of E_x , E_y and E_z as given by (3.21) are taken only the terms for which $m = \nu$, $n = 0$ remain. Furthermore, since the first power of l is a factor of the terms remaining in E_x and E_z and since W and $b(k, l)$ are even functions of l, it may be shown that the average values of E_x and E_z vanish. This is to be expected on physical grounds.

The average value of E_y is

$$(3.26) \qquad \begin{aligned} \langle E_y\rangle &= 2i \exp\{-i\beta\alpha x\} \sin \beta\gamma z \\ &\quad + 2\beta\gamma E(\nu, 0, z) \sum_{kl} \left[\frac{a^2l^2}{b(k, l)} + b(k, l)\right] \frac{\pi^2}{L^2} W(ak - a\nu, l) \\ &\to \exp\{-i\beta(\alpha x - \gamma z)\} \\ &\quad - \exp\{-i\beta(\alpha x + \gamma z)\}\left\{1 - 2\beta \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} ds \left[\frac{\gamma s^2}{b} + \gamma b\right] \frac{W(r - \beta\alpha, s)}{4}\right\} \end{aligned}$$

where we have used $a\nu = \beta\alpha$ and have set

$$r = ak = 2\pi k/L, \qquad s = al$$

$$(3.27) \qquad b = \begin{cases} [\beta^2 - r^2 - s^2]^{1/2}, & \beta^2 > r^2 + s^2 \\ -i[r^2 + s^2 - \beta^2]^{1/2}, & \beta^2 < r^2 + s^2. \end{cases}$$

In going from the summation to the integration we have assumed L to approach infinity just as in (2.5). By setting

$$p = r - \beta\alpha, \qquad q = s$$

$$(3.28) \qquad b = \begin{cases} [\beta^2 - (p + \beta\alpha)^2 - q^2]^{1/2} \\ -i[(p + \beta\alpha)^2 + q^2 - \beta^2]^{1/2} \end{cases}$$

the last term in (3.26) may be written as

$$(3.29) \qquad 2\beta \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq \left[\frac{\gamma q^2}{b} + \gamma b\right] \frac{W(p, q)}{4}.$$

The coefficient of $\exp\{-i\beta(\alpha x + \gamma z)\}$ in (3.26) represents the average value of the reflection coefficient and hence (3.29) represents the change in the reflection coefficient produced by the roughness.

The leading term in the mean square value of the fluctuation of E_y about the value it has in the absence of roughness is

$$\langle | E_y - 2i \exp\{-i\beta\alpha x\} \sin \beta\gamma z |^2 \rangle$$

$$= 4\beta^2\gamma^2 \sum_{mnkl} E^*(m, n, z)E(k, l, z)\langle P^*(m - \nu, n)P(k - \nu, l)\rangle$$

$$(3.30) \qquad = 4\beta^2\gamma^2 \sum_{kl} \exp\{-z\varphi(k, l)\}\pi^2 W(ak - a\nu, al)/L^2$$

$$\to 4\beta^2\gamma^2 \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} ds\, e^{-z\varphi} W(r - \beta\alpha, s)/4$$

$$= 4\beta^2\gamma^2 \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, e^{-z\varphi} W(p, q)/4$$

where

$$\varphi(k, l) = ib(k, l) - ib^*(k, l)$$

$$= \text{Imaginary part of } -2b(k, l)$$

$$(3.31) \qquad = \begin{cases} 0, & k^2 + l^2 < \beta^2/a^2 \\ 2[a^2k^2 + a^2l^2 - \beta^2]^{1/2}, & k^2 + l^2 > \beta^2/a^2 \end{cases}$$

and $\varphi = 0$ when $r^2 + s^2 < \beta^2$ or $(p + \beta\alpha)^2 + q^2 < \beta^2$ and

$$\varphi = 2[r^2 + s^2 - \beta^2]^{1/2} = 2[(p + \beta\alpha)^2 + q^2 - \beta^2]^{1/2} \tag{3.32}$$

when the inequalities are reversed. It is interesting to note that the average value of

$$E_y - 2i \exp\{-i\beta\alpha x\} \sin \beta\gamma z$$

is $O(\beta^2 f^2)$ (this is indicated by (3.26) and (3.29) since the double integral of $W(p, q)$ gives $\langle f^2(x, y)\rangle$) while the rms value of its modulus, as obtained by the square root of (3.30), is $O(\beta f)$.

When the procedure used to derive (3.30) is applied to the $O(f)$ terms in the expressions for E_z and E_x in (3.21) we obtain

$$\begin{aligned} \langle | E_z |^2\rangle &\to 4\beta^2\gamma^2 \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq \frac{e^{-z\varphi} q^2 W(p, q)}{4 | b |^2} \\ \langle | E_x |^2\rangle &= 0. \end{aligned} \tag{3.33}$$

where, from (3.28),

$$| b |^2 = | \beta^2 - (p + \beta\alpha)^2 - q^2 |. \tag{3.34}$$

Here we encounter trouble because the denominator may become zero. If $W(p, q)$ is continuous and not zero on the circle $| b |^2 = 0$ in the p, q plane, the double integral in (3.33) diverges logarithmically. This difficulty may be overcome in several ways. If the reflecting surface is not perfectly conducting a convergent double integral analogous to (3.33) may be obtained from the expression for $C_{mn}^{(1)}$ given by (7.21). When the conductivity g of the reflecting surface is large, but not infinite, equation (7.26) shows that the b occurring in the $| b |^2$ of the denominator of (3.33) should be replaced by $b + i\beta^2/\tau$ where τ is the intrinsic propagation constant of the reflecting material: $\tau \approx (i\omega\mu g)^{1/2}$, it being assumed that the permeability μ ($\mu = 4\pi \times 10^{-7}$ henries/meter for free space) is the same for the reflecting material as for the region $z > f(x, y)$. Since b is purely real or imaginary, it is seen that the new denominator never vanishes. Another method of meeting the difficulty is to assume $f(x, y) \equiv 0$ outside a square of side L instead of taking it to be a periodic function. The integral will converge as long as L remains finite.

Equations (3.30) and (3.33) show that the first approximation to the scattered field vanishes as z approaches infinity if $W(p, q)$ is zero for the region inside the circle $(p + \beta\alpha)^2 + q^2 = \beta^2$ where φ is zero, i.e., if the average distance between the hills is rather small compared to a wavelength. This means that the reflection in this case is perfect (the modulus of the average reflection coefficient being unity).

Incidentally, as Rayleigh has pointed out, the reflection from a simple sine wave surface will be perfect if the period of the sine wave is small enough.

In order to see this from our analysis suppose the equation of the surface to be $z = 2P \cos (m_1 ax + n_1 ay)$ so that all of the $P(m, n)$'s are zero except

$$P(m_1, n_1) = P(-m_1, -n_1) = P = \text{real}.$$

The only non-vanishing $O(f)$ terms in (3.21) are the two given by $m = \nu \pm m_1$, $n = \pm n_1$ (e.g., the upper signs go together), and the only non-vanishing $O(f^2)$ terms are the three given by $m = \nu \pm 2m_1$, $n = \pm 2n_1$ and $m = \nu$, $n = 0$. It follows that if m_1 and n_1 are such that

$$(|\nu| - |m_1|)^2 + n_1^2 > \beta^2/a^2 \tag{3.35}$$

the only term which can possibly correspond to a scattered wave is the one given by $m = \nu$, $n = 0$ (remember that $|\nu| < \beta/a$) because all of the others correspond to surface waves which carry no energy away from the surface. Since the $m = \nu$, $n = 0$ term corresponds to a wave traveling in the same direction as the main reflected wave it cannot be regarded as scattering. All it can do is change the phase of the reflection coefficient. Our work doesn't go beyond $O(f^2)$ terms but it doesn't seem likely that the higher order approximations will bring in any terms which can be interpreted as scattering.

However, the situation is quite different if the surface consists of the sum of two (or more) rapidly varying sine waves whose "interference pattern" has a period long enough to produce scattering. For example, let the surface be

$$z = 2P_1 \cos (m_1 ax + n_1 ay) + 2P_2 \cos (m_2 ax + n_2 ay) \tag{3.36}$$

where m_1, n_1 satisfy (3.35) and m_2, n_2 satisfy a similar inequality. An examination of (3.21) and the definition of $Q(m, n, k, l)$ shows that the $O(f^2)$ terms which might produce scattering are the two for which $m = \nu \pm (m_1 - m_2)$, $n = \pm(n_1 - n_2)$. At least one of these is certain to produce scattering if

$$(|\nu| - |m_1 - m_2|)^2 + (n_1 - n_2)^2 < \beta^2/a^2. \tag{3.37}$$

because it would correspond to a wave for which $b(m, n)$ is real and hence would carry energy away from the surface in a direction different from that of the main reflected wave. Even if (3.37) were not satisfied there is a possibility of higher order terms corresponding to scattering.

If we now consider the case of the rough surface with the above examples in mind we see that although the reflection may sometimes be perfect to a first approximation, the $O(f^2)$ terms in (3.21) give rise to a scattered field (somewhat similar to the Rayleigh scattering produced by small particles) which does not vanish as z becomes large. In order to study mean square values involving the $O(f^2)$ terms it is necessary to deal with averages of expressions containing the product of four $P(m, n)$'s. Since the results appear to be rather complicated, we shall not go farther than to state the following result which may be applied to our problem when P_n is replaced by $P(m, n)$ and the summation taken with respect to m, n instead of n, and likewise for k, n', k'.

Let P_0 be real. Let P_0 and the real and imaginary parts of $P_1, P_2, \cdots$

be independent random variables with average value zero. Let the real and imaginary parts of P_n, $n > 0$, have the same mean square value so that $\langle P_n^2 \rangle = 0$ unless $n = 0$, and define P_{-n} as the conjugate complex of P_n so that

$$\langle P_{-n} P_n \rangle = \langle | P_n |^2 \rangle = \langle | P_n^2 | \rangle,$$

$$\langle P_m P_n \rangle = 0 \qquad \text{if} \qquad m \neq -n.$$

If $F(n, k, n', k')$ denotes an arbitrary function of n, k, n', k' it can be shown that, if the summations run from $-\infty$ to $+\infty$,

$$\sum_{nkn'k'} F(n, k, n', k') \langle P_n P_k P_{n'} P_{k'} \rangle$$

$$= \sum_{nk} [F(n, -n, k, -k) + F(n, k, -n, -k) + F(n, k, -k, -n)]$$

$$\cdot \langle | P_n{}^2 | \rangle \langle | P_k^2 | \rangle \tag{3.38}$$

$$+ \sum_n [F(n, -n, n, -n) + F(n, n, -n, -n) + F(n, -n, -n, n)]$$

$$\cdot [\langle | P_n^4 | \rangle - 2(\langle | P_n^2 | \rangle)^2]$$

$$+ F(0, 0, 0, 0)[3(\langle P_0^2 \rangle)^2 - 2\langle P_0^4 \rangle].$$

One method of establishing this result is to break the four-fold summation into the subgroups for which (1) $k \neq n$, $k \neq -n$, (2) $k = n$, $n \neq 0$, (3) $k = -n$, $n \neq 0$, (4) $k = 0$, $n = 0$. The terms which have averages different from zero in subgroup (1) are those for which (1a) $n' = -n$, $k' = -k$, (1b) $n' = -k$, $k' = -n$. Likewise for the other groups we have (2a) $n' = k' = -n$, (3a) $n' = -n$, $k' = n$, (3b) $n' = n$, $k' = -n$, (3c) $n' = -k'$ but $n' \neq \pm n$, (4a) $n' = k' = 0$, (4b) $n' = -k'$, $n' \neq 0$.

When, as in the case of the rough surface, the surface $z = f(x, y)$ has many Fourier components of the same order of magnitude, the only term of importance on the right hand side of (3.38) is the double summation over n and k. This term goes into a fourfold integral involving the product of two $W(p, q)$ functions.

4. *Incident Wave Vertically Polarized*

In this section we assume the electric intensity of the field to be, in the absence of roughness,

$$E_x^a = 2i\gamma \exp\{-i\beta\alpha x\} \sin \beta\gamma z, \qquad E_y^a = 0$$
$$E_z^a = 2\alpha \exp\{-i\beta\alpha x\} \cos \beta\gamma z \tag{4.1}$$

where the symbols have the same meaning as in Section 3. In particular,

$\beta = 2\pi/\lambda$, and $(\alpha, 0, -\gamma)$, $(\alpha, 0, \gamma)$ are the direction cosines of the incident and reflected rays, respectively. A procedure similar to that used to obtain (3.21) leads to the following expressions, accurate to $O(f^2)$ terms, for the electric intensity in the presence of the slightly rough surface.

$$E_x = E_x^a + 2 \sum_{mn} E(m, n, z)[i(\alpha am - \beta)P(m - \nu, n)$$

$$+ \sum_{kl} \{a^2(m - k)(\nu - k)\beta + (\beta - \alpha am)b^2(k, l)\}Q(m, n, k, l)]$$

$$E_y = 2a \sum_{mn} E(m, n, z)[i\alpha nP(m - \nu, n)$$

$$+ \sum_{kl} \{a(n - l)(\nu - k)\beta - \alpha nb^2(k, l)\}Q(m, n, k, l)]$$

(4.2)
$$E_z = E_z^a + 2 \sum_{mn} [E(m, n, z)/b(m, n)]$$

$$\cdot \Big[i\{a(m - \nu)\beta + \alpha b^2(m, n)\}P(m - \nu, n)$$

$$+ \sum_{kl} \{a^3(k - \nu)(m^2 + n^2 - mk - nl)\beta$$

$$+ a[\alpha a(m^2 + n^2) - m\beta]b^2(k, l)\}Q(m, n, k, l) \Big].$$

In these equations $E(m, n, z)$ is the exponential function of x, y, z defined by (3.2) and $Q(m, n, k, l)$ is the function (3.22) containing the product of two P's.

The average electric intensity of the reflected wave is in the direction specified by the direction cosines $(-\gamma, 0, \alpha)$. The corresponding wave function approaches, as $L \to \infty$, $E(\nu, 0, z)$ multiplied by

(4.3)
$$1 - 2\beta \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} ds \left[\frac{(r - \beta\alpha)^2}{\gamma b} + \gamma b\right] W(r - \beta\alpha, s)/4$$

$$= 1 - 2\beta \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq \left[\frac{p^2}{\gamma b} + \gamma b\right] W(p, q)/4$$

where b is defined as a function of r, s and p, q by equations (3.27) and (3.28). The derivation of (4.3) is similar to that of its analogue in (3.26): the expressions (4.2) are averaged, the reflected wave picked out, and the square root of the sum of the squares of its x, y, z components taken (the average y component turns out to be zero).

An idea of how the field components vary about their values in the absence of roughness may be obtained from the following analogues of (3.30) and (3.33).

$$\langle | E_x - E_x^a |^2 \rangle = 4 \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, e^{-z\varphi}(\alpha p - \beta\gamma^2)^2 W(p, q)/4$$

$$\langle | E_y |^2 \rangle = 4 \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, e^{-z\varphi}\alpha^2 q^2 W(p, q)/4 \tag{4.4}$$

$$\langle | E_z - E_z^a |^2 \rangle = 4 \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, e^{-z\varphi}(p\beta + \alpha b^2)^2 W(p, q)/4 \, | b |^2.$$

Here, as in (3.33), the last integral may not converge on the circle $b = 0$. It was pointed out that this difficulty can be overcome in the case of horizontal polarization by considering the electrical properties of the reflecting surface, and the same is probably also true for the case of vertical polarization.

The analogue of (3.23) is

$$\text{Real Part of } 2\beta A_{\nu 0} = \sum [| A_{mn}^2 | + | B_{mn}^2 | + | C_{mn}^2 |] b(m, n) \tag{4.5}$$

where the summation extends only over those values of m and n for which $b(m, n)$ is real $(m^2 + n^2 \leq \beta^2/a^2)$ and A_{mn} etc. are defined by

$$E_x = E_x^a + \sum A_{mn} E(m, n, z)$$

$$E_y = \sum B_{mn} E(m, n, z)$$

$$E_z = E_z^a + \sum C_{mn} E(m, n, z).$$

Equation (4.5) and $C_{\nu 0} = -\alpha A_{\nu 0}/\gamma$, which follows from the divergence relation (3.6), give a partial check on equations (4.2).

5. *Special Cases*

Suppose that the roughness spectrum, $W(p, q)$ is zero except for a small region around $p = 0$, $q = 0$. In this case the average distance between the hills of the surface is large compared to the wavelength of the incident radiation. The function b defined by (3.28) differs but little from its value at $p = q = 0$, namely $\beta\gamma$, and (3.29) becomes

$$2\beta \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, \gamma\beta\gamma W(p, q)/4 = 2\beta^2\gamma^2 \langle f^2(x, y) \rangle. \tag{5.1}$$

Here we have used expression (2.5) for the mean square value of $f(x, y)$ and have assumed $\gamma q^2/b$ in (3.29) to be negligibly small in the region where $W(p, q)$ is different from zero. The average value of the reflection coefficient for horizontal polarization now becomes

$$1 - 2\beta^2\gamma^2 \langle f^2(x, y) \rangle. \tag{5.2}$$

A similar treatment of (4.3) which involves the neglect of $p^2/\gamma b$ shows

that (5.2) also holds for the case of vertical polarization. It is interesting to note that (5.2) agrees with the first two terms in the expansion of a result obtained by W. S. Ament in which βf is not required to be small, namely, that the roughness reduces the amplitude of the average reflected waves by the factor $\exp\{-2\beta^2\gamma^2\langle f^2(x, y)\rangle\}$. As pointed out in the introduction, this agreement is all that the approximate nature of our results will allow.

When $W(p, q)$ differs from zero only in the region around $p = 0$, $q = 0$, equations (3.30) and (3.33) show that for horizontal polarization

$$\begin{aligned} &\langle | E_y - 2i \exp\{-i\beta\alpha x\} \sin \beta\gamma z |^2\rangle = 4\beta^2\gamma^2\langle f^2(x, y)\rangle, \\ &\langle | E_z |^2\rangle = 4\langle f_y^2(x, y)\rangle \end{aligned} \tag{5.3}$$

where $f_y(x, y) = \partial f(x, y)/\partial y$; equations (4.4) show that for vertical polarization

$$\begin{aligned} &\langle | E_x - E_x^a |^2\rangle = 4\beta^2\gamma^4\langle f^2(x, y)\rangle \\ &\langle | E_y |^2\rangle = 4\alpha^2\langle f_y^2(x, y)\rangle \\ &\langle | E_z - E_z^a |^2\rangle = 4\beta^2\gamma^2\alpha^2\langle f^2(x, y)\rangle. \end{aligned} \tag{5.4}$$

Suppose now that $W(p, q)$ is such as to make the terms $\gamma q^2/b$ and $p^2/\gamma b$, which were neglected above, the dominant terms in the integrands of (3.29) and (4.3). The average distance between hills of the corresponding surface will now be small compared to a wavelength. The magnitudes of the average reflection coefficients are then approximately

$$\begin{aligned} &1 - 2\beta \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, \gamma q^2 W(p, q)/4b = 1 - \gamma s_p \\ &1 - 2\beta \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, p^2 W(p, q)/4\gamma b = 1 - \frac{s_h}{\gamma} \end{aligned} \tag{5.5}$$

for horizontal and vertical polarizations, respectively. Here s_p and s_h stand for small quantities, and γ is the cosine of the angle between the z-axis and the reflected ray. The remarkable thing about the reflection coefficients (5.5) is that they depend on γ in the same way as do the corresponding reflection coefficients, computed from Fresnel's formulas, for a good, but not perfect, plane conductor.

For vertical incidence $\gamma = 1$, $\alpha = 0$ and the two expressions given by (5.5) reduce to essentially the same thing, the q^2 in the first expression (where the incident E is parallel to the y-axis) goes over to the p^2 in the second expression (where the incident E is parallel to the x-axis) because of the difference in the assumed incident waves.

6. *Propagation along Surface*

As the condition of grazing incidence is approached γ approaches zero, and expression (4.3) giving the average reflection coefficient for vertical polarization breaks down. In this case a modification of the method used to study reflection may be used to obtain a solution corresponding to a wave guided by the surface. A solution of this sort is to be expected since it has been known for some time that a corrugated or slotted surface will support a typical "surface wave" in which the field decreases exponentially with distance from the surface.

To start with, we take the perfectly conducting surface to be

$$z = 2P \cos sx = f \tag{6.1}$$

which shows that f is now merely a function of x. Guided by the known properties of surface waves, we assume that there exists a wave in which the electric intensity is predominantly in the z direction (approximately normal to the surface) and that there is also a small component of E in the x direction (in the direction of propagation). We also tacitly assume that the velocity of propagation of the wave does not differ much from that of a wave traveling freely in the medium above the surface; i.e., if the propagation of the principal part of the wave is described by $\exp\{i\omega t - ihsx\}$ then hs approaches $\beta = 2\pi/\lambda$ as the amplitude P of the corrugations approaches zero.

When we attempt to express our assumptions as equations some experimentation suggests the forms

$$\begin{aligned} E_x &= \sum_m A_m E(h + m, z), \qquad E_y = 0 \\ E_z &= E(h, z) + \sum_m C_m E(h + m, z) \end{aligned} \tag{6.2}$$

where the summations with respect to m extend over all integers from $-\infty$ to ∞ and A_m, C_m are small quantities which approach zero with P. In order to fix the amplitudes of the various components C_0 is taken to be zero so that there is no term corresponding to $E(h, z)$ in the summation part of E_z. Here

$$\begin{aligned} E(h + m, z) &= \exp\{-i(h + m)sx - ib(h + m)z\} \\ b(h + m) &= \begin{cases} [\beta^2 - (h + m)^2 s^2]^{1/2}, & \beta^2 > (h + m)^2 s^2 \\ -i[(h + m)^2 s^2 - \beta^2]^{1/2}, & \beta^2 < (h + m)^2 s^2 \end{cases} \end{aligned} \tag{6.3}$$

so that the components (6.2) satisfy the wave equation. Since the difference between β and hs is assumed to be small it follows that $b(h)$ is to be regarded as small.

Since we do not intend to carry our approximations beyond $O(\beta^2 f^2)$ we may use the first of the boundary conditions (3.10) which, for our surface (6.1), becomes

$$E_x = N_x E_z = -f_x E_z = 2Ps \sin sx E_z \tag{6.4}$$
$$= Ps(-ie^{isx} + ie^{-isx})E_z .$$

This relation must be satisfied at $z = 2P \cos sx = f$ to within an accuracy of $O(P^2)$.

Upon substituting the assumed expressions (6.2) for E_x and E_z in the boundary condition (6.4), using relations of the form

$$A = A_m^{(1)} + A_m^{(2)} + \cdots$$

$$E(h + m, f) = E(h + m, 0)[1 - ib(h + m)f + \cdots]$$

$$E(h + m, 0)f = PE(h + m - 1, 0) + PE(h + m + 1, 0)$$

in the same manner as in the reflection problem, and equating first order terms we see that

$$A_1^{(1)} = Psi, \qquad A_{-1}^{(1)} = -Psi \tag{6.5}$$
$$A_m^{(1)} = 0 \qquad \text{if} \qquad m \neq 1 \text{ or } -1.$$

The divergence relation div $E = 0$ gives

$$(h + m)sA_m + b(h + m)C_m = 0, \qquad m \neq 0 \tag{6.6}$$
$$hsA_0 + b(h) = 0, \qquad m = 0.$$

Since $A_0^{(1)}$ is zero, $b(h)$ is smaller than a first order term (it will be shown later to be $O(P^2)$). From the first of equations (6.6) it follows that

$$C_1^{(1)} = -(h + 1)sA_1^{(1)}/b(h + 1), \qquad C_{-1}^{(1)} = -(h - 1)sA_{-1}^{(1)}/b(h - 1) \tag{6.7}$$
$$C_m^{(1)} = 0 \qquad \text{if} \qquad m \neq 1, 0 \text{ or } -1.$$

Equating the second order terms in (6.4), and using (6.5) and (6.7) gives

$$A_0^{(2)} = iP[b(h + 1)A_1^{(1)} - sC_1^{(1)} + b(h - 1)A_{-1}^{(1)} + sC_{-1}^{(1)}]$$
$$= P^2 s\left[\frac{hs^2 - \beta^2 + h^2s^2}{b(h + 1)} + \frac{hs^2 + \beta^2 - h^2s^2}{b(h - 1)}\right]$$

$$A_2^{(2)} = iP[b(h + 1)A_1^{(1)} + sC_1^{(1)}] \tag{6.8}$$

$$A_{-2}^{(2)} = iP[b(h - 1)A_{-1}^{(1)} - sC_{-1}^{(1)}]$$

$$A_m^{(2)} = 0 \qquad \text{if} \qquad m \neq 0, 2 \text{ or } -2.$$

The expression for $A_0^{(2)}$ is of particular importance because when it is combined

with the second of equations (6.6), which we write as $A_0^{(2)} = -b(h)/hs$, we obtain an equation which may be solved for the propagation constant hs in the x direction:

$$-b(h) = P^2hs^2\left[\frac{hs^2 - \beta^2 + h^2s^2}{b(h+1)} + \frac{hs^2 + \beta^2 - h^2s^2}{b(h-1)}\right]. \tag{6.9}$$

This expression shows that $b(h)$ is $O(P^2)$ and therefore, when P is small in accordance with our assumptions, hs is nearly equal to β. Replacing hs by β in (6.9) gives

$$b(h) \approx -iP^2\beta^2s[(1 + 2\beta/s)^{-1/2} + (1 - 2\beta/s)^{-1/2}] \tag{6.10}$$

which shows that if $s > 2\beta$, $b(h)$ is negative imaginary and $E(h, z)$ decreases exponentially with increasing z. Thus in this case we have a true surface wave.

When s is much greater than β so that the surface has many corrugations in one wavelength of the electromagnetic wave, we get from (6.10)

$$b(h) \approx -2iP^2\beta^2s$$

$$hs \approx \beta + 2P^4\beta^3s^2$$

and the principal part of the field is the surface wave

$$E_z \approx \exp\{-i\beta(1 + 2P^4\beta^2s^2)x - 2P^2\beta^2sz\} \tag{6.11}$$

which travels a little more slowly than a free wave.

The same type of analysis may be used to investigate the surface wave which is guided by the more general rough surface described in Section 2. We assume

$$\begin{aligned} E_x &= \sum_{mn} A_{mn}E(m+h, n, z) \\ E_y &= \sum_{mn} B_{mn}E(m+h, n, z) \\ E_z &= E(h, 0, z) + \sum_{mn} C_{mn}E(m+h, n, z) \end{aligned} \tag{6.12}$$

where the summations extend over all integral values of m and n between plus and minus infinity, $C_{00} = 0$, and $E(m + h, n, z)$ is defined by (3.2) and (3.3) with m replaced by $m + h$. The situation is somewhat similar to putting $\gamma = 0$, $\alpha = 1$ in the vertical polarization case of reflection. The boundary conditions are

$$E_x = -f_xE_z\,, \qquad E_y = -f_yE_z \tag{6.13}$$

and these, together with the condition div $E = 0$:

$$
(6.14) \quad \begin{aligned} & a(m+h)A_{mn} + anB_{mn} + b(h+m, n)C_{mn} = 0, \qquad m, n \neq 0 \\ & ahA_{00} + b(h, 0) = 0, \qquad m = n = 0 \end{aligned}
$$

lead to the expressions

$$
(6.15) \quad \begin{aligned} & A^{(1)}_{mn} = iamP(m, n), \qquad B^{(1)}_{mn} = ianP(m, n) \\ & C^{(1)}_{mn} = -ia^2[m^2 + mh + n^2]P(m, n)/b(h + m, n) \end{aligned}
$$

for the $O(f)$ terms in the coefficients. The $O(f^2)$ terms in A_{mn} and B_{mn} are

$$
(6.16) \quad \begin{aligned} A^{(2)}_{mn} &= \sum_{kl} i[a(m-k)C^{(1)}_{kl} + b(h+k, l)A^{(1)}_{kl}]P(m-k, n-l) \\ B^{(2)}_{mn} &= \sum_{kl} i[a(n-l)C^{(1)}_{kl} + b(h+k, l)B^{(1)}_{kl}]P(m-k, n-l). \end{aligned}
$$

Since $A^{(1)}_{00}$ is zero, from (6.15), $A^{(2)}_{00}$ is given by the second of equations (6.14). Equating this to the value of $A^{(2)}_{00}$ given by (6.16) leads to

$$
(6.17) \quad b(h, 0) = \sum_{kl} a^2hk(\beta^2 - a^2h^2 - a^2hk) \mid P(k, l) \mid^2 / b(h + k, l).
$$

As the roughness decreases, $b(h, 0)$ approaches zero and ha approaches β and we have

$$
b(h, 0) \approx -\sum_{kl} a^2\beta^2k^2 \mid P(k, l) \mid^2 / b(k + \beta/a, l).
$$

When this is averaged over the universe of rough surfaces mentioned in Section 2 and when (2.3) is used we obtain, upon letting L approach infinity,

$$
(6.18) \quad b(h, 0) \approx -\int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, \beta^2p^2W(p, q)/4b_1
$$

where b_1 is the function of p and q obtained by setting $\alpha = 1$ in expression (3.28) for b:

$$
(6.19) \quad b_1 = \begin{cases} [\beta^2 - (p + \beta)^2 - q^2]^{1/2} \\ -i[(p + \beta)^2 + q^2 - \beta^2]^{1/2}. \end{cases}
$$

The principal part of the surface wave is

$$
E_s \approx \exp\{-iahx - ib(h, 0)z\}
$$

which leads us to introduce $B = ib(h, 0)$ so that

$$
B^2 = -\beta^2 + a^2h^2 \approx (ah - \beta)2\beta
$$

$$
ah \approx \beta + B^2/2\beta.
$$

We may therefore summarize our result by saying that the principal part of the surface wave corresponding to the general (slightly rough) surface of Section 2 is

$$E_s \approx \exp\{-i\beta(1 + B^2/2\beta^2)x - Bz\}$$

where

$$B = B_r + iB_i = -i\beta^2 \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, p^2W(p, q)/4b_1$$

and the attenuation in the x direction is $-B_rB_i/\beta$ (nepers/meter). The definition (6.19) of b_1 shows that B_i is never positive. It also shows that if $W(p, q)$ is zero where b_1 is real, namely inside the circle of radius β centered at $p = -\beta$, $q = 0$, B_i is zero and there is no attenuation. This corresponds to the case where the hills of the surface are close together and is in agreement with the view that the guiding action of the surface is due to rapidly undulating components of $z = f(x, y)$ while the attenuation is due to the scattering produced by the more slowly varying components. It should be remembered that (6.18) is only an approximate expression for $b(h, 0)$. It seems probable that more accurate expressions would show an attenuation even if $W(p, q)$ were zero in the circle mentioned above because this is no guarantee that $A_{mn}^{(2)}$ and $B_{mn}^{(2)}$ given by (6.16) will vanish for values of m and n which correspond to waves carrying energy away from the surface. Thus it appears that even though the surfaces $z = P \cos sx$ and $z = Q \cos tx$ can carry surface waves without attentuation when $s > 2\beta$ and $t > 2\beta$, the same is not true of the surface $z = P \cos sx + Q \cos tx$ if, for example, $s - t$ were almost equal to β. The situation is somewhat similar to the one encountered in the discussion of reflection from the surface (3.36).

7. *Reflection from Wavy Interface between Two Media—Horizontal Polarization*

Let the interface coincide approximately with the plane $z = 0$ and let the propagation constants σ and τ of the upper ($z > 0$) and lower media, respectively, be given by

$$\sigma = i\omega(\mu\epsilon_0)^{1/2} = i\beta, \qquad \tau = \sigma(\epsilon_r + g/i\omega\epsilon_0)^{1/2}. \tag{7.1}$$

Here we have assumed that both media have the same permeability μ and that the ratio of their dielectric constants is ϵ_r. g is the conductivity of the lower medium and ϵ_0 the dielectric constant of the upper medium. The upper medium is non-conducting. For free space $\mu = 1.257 \times 10^{-6}$ henry/meter and $\epsilon_0 = 8.854 \times 10^{-12}$ farad/meter.

If the interface coincided exactly with the plane $z = 0$ the electric intensity for horizontal polarization would be

$$E_y = E^+ \equiv \exp\{-\sigma\alpha x\}(\exp\{\sigma\gamma z\} + R\exp\{-\sigma\gamma z\}), \qquad z > 0$$

$$E_y = E^- \equiv T\exp\{-\sigma\alpha x + \tau\gamma' z\}, \qquad z < 0$$

(7.2)

$$\tau\alpha' = \sigma\alpha = iav, \qquad \gamma' = (1 - \alpha'^2)^{1/2}$$

$$R = \frac{1 - \dfrac{\tau\gamma'}{\sigma\gamma}}{1 + \dfrac{\tau\gamma'}{\sigma\gamma}}, \qquad T = \frac{2}{1 + \dfrac{\tau\gamma'}{\sigma\gamma}}.$$

As before, $\alpha = \sin\theta$ and $\gamma = \cos\theta$ where θ is the angle between the z-axis and the reflected ray.

When the equation of the separating surface is $z = f(x, y) \equiv f$ we assume the electric intensity to be

$$E_x = \begin{cases} \sum A_{mn}E(m, n, z) & \text{for} \quad z > f \\ \sum G_{mn}F(m, n, z) & \text{for} \quad z < f \end{cases}$$

(7.3)

$$E_y = \begin{cases} E^+ + \sum B_{mn}E(m, n, z) & \text{for} \quad z > f \\ E^- + \sum H_{mn}F(m, n, z) & \text{for} \quad z < f \end{cases}$$

$$E_z = \begin{cases} \sum C_{mn}E(m, n, z) & \text{for} \quad z > f \\ \sum I_{mn}F(m, n, z) & \text{for} \quad z < f \end{cases}$$

where E^+, E^- are given by (7.2) (with the dividing surface $z = 0$ replaced by $z = f(x, y)$) and

$$E(m, n, z) = \exp\{-ia(mx + ny) - ib(m, n)z\}$$

$$F(m, n, z) = \exp\{-ia(mx + ny) + ic(m, n)z\}$$

(7.4)

$$ib(m, n) = [\sigma^2 + a^2(m^2 + n^2)]^{1/2} \qquad a = 2\pi/L$$

$$ic(m, n) = [\tau^2 + a^2(m^2 + n^2)]^{1/2}.$$

Here $b(m, n)$ is the same as the $b(m, n)$ defined by (3.3) and is either positive real or negative imaginary. The same would be true of $c(m, n)$ if the lower medium were non-conducting.

At $z = f$ we require the continuity of

$$E_x - N_x(N_xE_x + N_yE_y + N_zE_z)$$

(7.5)

$$E_y - N_y(N_xE_x + N_yE_y + N_zE_z)$$

and two other expressions obtained by substituting H (the magnetic intensity) for E. When we assume the components N_x, N_y of the normal to be small (so that $N_z \approx 1$), and also assume E_x and E_z to be small, (7.5) becomes

$$(7.6) \qquad \begin{gathered} E_x - N_x N_y E_y - N_x E_z \\ (1 - N_y^2) E_y - N_y E_z . \end{gathered}$$

The H conditions corresponding to (7.5) and the assumptions that N_x, N_y, E_x, E_y and their derivatives must be small tell us that the two expressions

$$(7.7) \qquad \begin{gathered} \frac{\partial E_z}{\partial y} - (1 - N_x^2) \frac{\partial E_y}{\partial z} - N_x \frac{\partial E_y}{\partial x} + N_x \frac{\partial E_x}{\partial y} \\ \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} + N_y N_x \frac{\partial E_y}{\partial z} - N_y \frac{\partial E_y}{\partial x} + N_y \frac{\partial E_x}{\partial y} \end{gathered}$$

must be continuous at $z = f$. Here we have made use of the assumption that the two media have the same permeability, and have neglected $O(f^3)$ terms. The terms $N_x^2\ \partial E_y/\partial z$ and $N_y N_x\ \partial E_y/\partial z$ may be omitted from (7.7) since the first of the two relations

$$(7.8) \qquad \begin{gathered} \sigma\gamma(1 - R) = T\tau\gamma' \\ 1 + R = T \end{gathered}$$

ensures the continuity (out to $O(f^2)$) of the terms in question. In the same way, the second of relations (7.8) enables us to omit $N_x N_y E_y$ and $N_y^2 E_y$ from (7.6).

When the assumed expressions (7.3) for the electric intensity are set in the boundary conditions (7.6) and (7.7), as just amended, the terms arising from E^+ and E^- can be simplified by using (7.8). For example, in the second of equations (7.6) these terms are

$$(7.9) \qquad \begin{aligned} &\exp\{-\sigma\alpha x\}(\exp\{\sigma\gamma f\} + R\exp\{-\sigma\gamma f\} - T\exp\{\tau\gamma' f\}) \\ &\quad = \exp\{-\sigma\alpha x\} f^2 U + O(f^3) \end{aligned}$$

where

$$(7.10) \qquad U = T(\sigma^2 - \tau^2)/2.$$

After similar reductions are made in (7.7), the four relations arising from (7.6) and (7.7) may be written as

$$\sum \{[A_{mn} + f_x C_{mn}]E(m, n, f) - [G_{mn} + f_x I_{mn}]F(m, n, f)\} = 0$$

$$\begin{aligned} \exp\{-\sigma\alpha x\} f^2 U + \sum \{&[B_{mn} + f_y C_{mn}]E(m, n, f) \\ &- [H_{mn} + f_y I_{mn}]F(m, n, f)\} = 0 \end{aligned}$$

$$(7.11) \qquad -\exp\{-\sigma\alpha x\} U[2f + \tau\gamma' f^2] + i \sum \{[-anC_{mn} + b(m, n)B_{mn}$$

$$- f_x amB_{mn} + f_x anA_{mn}]E(m, n, f)$$

$$- [-anI_{mn} - c(m, n)H_{mn} - f_x amH_{mn} + f_x anG_{mn}]F(m, n, f)\} = 0$$

$$\sum \{[-b(m, n)A_{mn} + amC_{mn} - f_y amB_{mn} + f_y anA_{mn}]E(m, n, f)$$

$$- [c(m, n)G_{mn} + amI_{mn} - f_y amH_{mn} + f_y anG_{mn}]F(m, n, f)\} = 0$$

where $O(f^3)$ terms are neglected.

We now assume α is such that $\sigma\alpha = ia\nu$ where ν is an integer. In order to separate the first and second order terms in (7.11) we write the various coefficients as $A_{mn}^{(1)} + A_{mn}^{(2)} + \cdots$, and so on, and use the approximate expressions

$$\begin{aligned} E(m, n, f) &= [1 - ib(m, n)f]E(m, n, 0) \\ F(m, n, f) &= [1 + ic(m, n)f]E(m, n, 0). \end{aligned} \tag{7.12}$$

By replacing $f \exp\{-ia\nu x\}$ by its Fourier series expansion (3.16) and proceeding as in Section 3 we find that the first order terms in (7.11) lead to

$$\begin{aligned} A_{mn}^{(1)} = G_{mn}^{(1)}, \qquad B_{mn}^{(1)} &= H_{mn}^{(1)} \\ id(m, n)B_{mn}^{(1)} - ian(C_{mn}^{(1)} - I_{mn}^{(1)}) &= 2UP(m - \nu, n) \\ -d(m, n)A_{mn}^{(1)} + am(C_{mn}^{(1)} - I_{mn}^{(1)}) &= 0 \end{aligned} \tag{7.13}$$

where

$$d(m, n) = b(m, n) + c(m, n). \tag{7.14}$$

The equations arising from the second order terms in the first two of equations (7.11) may be simplified with the help of equation (3.17), the relations (7.13) between the first order terms, and the expansion

$$\exp\{-ia\nu x\}f^2 = \sum P(k - \nu, l)P(m - k, n - l)E(m, n, 0) \tag{7.15}$$

where the summation on the right extends over all integral values of m, n, k, l from $-\infty$ to $+\infty$. In dealing with the last two equations of (7.11) we need the additional results

$$\begin{aligned} c^2(m, n) - b^2(m, n) &= \sigma^2 - \tau^2 \\ b(m, n)C_{mn}^{(1)} + c(m, n)I_{mn}^{(1)} &= 0 \end{aligned} \tag{7.16}$$

the first of which follows from the definitions of $c(m, n)$ and $b(m, n)$ and the second from subtraction of the first order terms in the two div $E = 0$ equations

$$\begin{aligned} amA_{mn} + anB_{mn} + b(m, n)C_{mn} &= 0 \\ amG_{mn} + anH_{mn} - c(m, n)I_{mn} &= 0. \end{aligned} \tag{7.17}$$

The results of this simplification are given by the equations

$$
(7.18)\qquad
\begin{aligned}
A^{(2)}_{mn} - G^{(2)}_{mn} &= h_1\\
B^{(2)}_{mn} - H^{(2)}_{mn} &= h_2\\
an(C^{(2)}_{mn} - I^{(2)}_{mn}) - b(m, n)B^{(2)}_{mn} - c(m, n)H^{(2)}_{mn} &= h_3\\
am(C^{(2)}_{mn} - I^{(2)}_{mn}) - b(m, n)A^{(2)}_{mn} - c(m, n)G^{(2)}_{mn} &= h_4
\end{aligned}
$$

where, taking the summations over k and l,

$$
(7.19)\qquad
\begin{aligned}
h_1 &= iam \sum (C^{(1)}_{kl} - I^{(1)}_{kl})P(m - k, n - l)\\
h_2 &= \sum [UP(k - \nu, l) + ian(C^{(1)}_{kl} - I^{(1)}_{kl})]P(m - k, n - l)\\
h_3 &= i \sum [U\tau\gamma' P(k - \nu, l) + (\sigma^2 - \tau^2)B^{(1)}_{kl}]P(m - k, n - l)\\
h_4 &= i(\sigma^2 - \tau^2) \sum A^{(1)}_{kl}P(m - k, n - l).
\end{aligned}
$$

Equations (7.13), (7.17) and (7.18) may now be used to obtain expressions, valid as far as $O(f^2)$, for the coefficients. From (7.16)

$$
(7.20)\qquad I^{(1)}_{mn} = -\frac{b(m, n)}{c(m, n)} C^{(1)}_{mn}, \qquad C^{(1)}_{mn} - I^{(1)}_{mn} = \frac{d(m, n)}{c(m, n)} C^{(1)}_{mn}
$$

and these relations enable us to derive the expressions

$$
(7.21)\qquad
\begin{aligned}
A^{(1)}_{mn} = G^{(1)}_{mn} &= \frac{i2Ua^2mnP(m - \nu, n)}{d(m, n)D_{mn}}\\
B^{(1)}_{mn} = H^{(1)}_{mn} &= \frac{i2UP(m - \nu, n)}{d(m, n)} \left[\frac{a^2n^2}{D_{mn}} - 1\right]\\
C^{(1)}_{mn} &= \frac{i2Uan\, c(m, n)P(m - \nu, n)}{d(m, n)D_{mn}}\\
C^{(1)}_{mn} - I^{(1)}_{mn} &= \frac{i2UanP(m - \nu, n)}{D_{mn}}
\end{aligned}
$$

where

$$
(7.22)\qquad D_{mn} = a^2(m^2 + n^2) + b(m, n)c(m, n).
$$

Explicit expressions for the h's are obtained when (7.21) is put in (7.19). The second order terms may be obtained from

$$
(7.23)\qquad
\begin{aligned}
dDA^{(2)}_{mn} &= a^2m^2bh_1 + (D - a^2m^2)(ch_1 - h_4) + a^2mn(bh_2 - ch_2 + h_3)\\
dDB^{(2)}_{mn} &= a^2n^2bh_2 + (D - a^2n^2)(ch_2 - h_3) + a^2mn(bh_1 - ch_1 + h_4)\\
dDC^{(2)}_{mn} &= \tau^2a(mh_1 + nh_2) + ca(mh_4 + nh_3)
\end{aligned}
$$

by dividing through by dD (where we have written d, b, c, and D for $d(m, n)$, $b(m, n)$, $c(m, n)$, and D_{mn}). These expressions are obtained from (7.18) and (7.17) (written out for the second order terms).

The manner in which these expressions approach the earlier expressions for the perfect conductor may be examined by letting the conductivity g approach infinity. From (7.1) we see that, since $\sigma = i\beta$, τ behaves like a large positive number multiplied by $i^{1/2}$. From equations (7.2, 4, 10, 14, and 22)

$$\alpha' = \sigma\alpha/\tau, \qquad \gamma' = 1 + O(\tau^{-2})$$

$$T = \frac{2\sigma\gamma}{\tau} + O(\tau^{-2}), \qquad U = -\sigma\gamma\tau + O(1)$$

$$c(m, n) = -i\tau + O(\tau^{-1}), \quad d(m, n) = -i\tau + b(m, n) + O(\tau^{-1}) \tag{7.24}$$

$$D_{mn} = -i\tau b(m, n) + a^2(m^2 + n^2) + O(\tau^{-1})$$

$$iU/d(m, n) = \sigma\gamma + O(1).$$

In the case of perfect conductivity studied in Sections 3 and 4, one source of annoyance was the appearance of $b(m, n)$ as a factor in certain denominators. Here the corresponding term is $-i\tau b(m, n)$ in D_{mn}. Since $b(m, n)$ may become small, or even vanish, we have retained the $a^2(m^2 + n^2)$ term in D_{mn}.

When τ becomes large equations (7.21) become

$$\begin{aligned} A_{mn}^{(1)} &= \frac{2\sigma\gamma a^2 mnP(m-\nu, n)}{a^2(m^2+n^2) - i\tau b(m, n)} \to 0 \\ B_{mn}^{(1)} &= 2\sigma\gamma P(m-\nu, n)\left[\frac{a^2n^2}{a^2(m^2+n^2) - i\tau b(m, n)} - 1\right] \\ &\to -2\sigma\gamma P(m-\nu, n) \\ C_{mn}^{(1)} &= \frac{2\sigma\gamma anP(m-\nu, n)}{b(m, n) + ia^2(m^2+n^2)/\tau}. \end{aligned} \tag{7.25}$$

When $b(m, n)$ is very small $a^2(m^2 + n^2)$ is nearly equal to $-\sigma^2$ and we may replace the denominator in $C_{mn}^{(1)}$ by

$$b(m, n) + ia^2(m^2 + n^2)/\tau \approx b(m, n) - i\sigma^2/\tau = b(m, n) + i\beta^2/\tau \tag{7.26}$$

which never vanishes since $b(m, n)$ is either positive real or negative imaginary. Thus the difficulty encountered in Section 3 (and, presumably, also that in Section 4) may be overcome by taking the electrical properties of the reflecting surface into account.

The average value of E_y in the upper medium, from which the average

value of the reflection coefficient may be obtained turns out to be the average value of

$$E^+ + B_{\nu 0}^{(2)}E(\nu, 0, z) = \exp\{-\sigma\alpha x\}[\exp\{\sigma\gamma z\} + \exp\{-\sigma\gamma z\}(R + B_{\nu 0}^{(2)})]$$

$$d(\nu, 0)B_{\nu 0}^{(2)} = c(\nu, 0)h_2 - h_3 \tag{7.27}$$

$$\langle B_{\nu 0}^{(2)}\rangle = \frac{2U}{d(\nu, 0)} \sum_{kl} \left\{\frac{(\sigma^2 - \tau^2)}{d(k, l)}\left[\frac{a^2 l^2}{D_{kl}} - 1\right] - i\tau\gamma'\right\} \frac{\pi^2 W(r - \alpha\beta, s)}{L^2}$$

where we have used the relations

$$ic(\nu, 0) = \tau\gamma', \qquad a\nu = \beta\alpha = \sigma\alpha/i$$

$$r = ak = 2\pi k/L, \qquad s = al.$$

When we let L approach infinity, the double summation may be replaced by a double integration in the usual way and we get, after some reduction,

$$\begin{aligned}\langle B_{\nu 0}^{(2)}\rangle = \frac{i2\sigma\gamma(\sigma^2 - \tau^2)}{(\tau\gamma' + \sigma\gamma)^2} \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} ds\, \frac{W(r - \beta\alpha, s)}{4} \Bigg[-i\tau\gamma' \\ + \frac{(\sigma^2 - \tau^2)}{c + b}\left(\frac{s^2}{r^2 + s^2 + bc} - 1\right)\Bigg]\end{aligned} \tag{7.28}$$

where c and b denote functions of r and s defined by

$$\begin{aligned} ic &= (\tau^2 + r^2 + s^2)^{1/2} \\ ib &= (\sigma^2 + r^2 + s^2)^{1/2} = i(\beta^2 - r^2 - s^2)^{1/2}. \end{aligned} \tag{7.29}$$

As g approaches infinity (7.28) should approach the value of its counterpart, given by the double integral in (3.26), which was obtained in Section 3 for reflection from a perfectly conducting but slightly rough surface. That this is the case may be verified with the help of expressions (7.24) which hold for large values of g.

Bibliography

1. Feinberg, E., *On the propagation of radio waves along an imperfect surface,* Journal of Physics (USSR), Volume 8, 1944, pp. 317–330; Volume 9, 1945, pp. 1–6; Volume 10, 1946, pp. 410–418.
2. Blake, L. V., *Reflection of radio waves from a rough sea,* Proceedings of the Institute of Radio Engineers, Volume 38, 1950, pp. 301–304.
3. Carlson, J. F., and Heins, A. E., *The reflection of an electromagnetic plane wave by an infinite set of plates,* Quarterly of Applied Mathematics, Volume 4, 1946, pp. 313–329; Volume 5, 1947, pp. 82–88.
4. Lord Rayleigh. *Theory of Sound,* Volume II, Macmillan, London, 1929, pp. 89–96.
5. Goldstein, H., Thesis, Massachusetts Institute of Technology, 1943.
6. Brillouin, L., *Wave guides for slow waves,* Journal of Applied Physics, Volume 19, 1948, pp. 1023–1041.
7. Twersky, V., *On the non-specular reflection of plane waves of sound,* Journal of the Acoustical Society of America, Volume 22, 1950, pp. 539–546. In this article, published after the present paper was written, Dr. Twersky has given an account of some of his work.

The Theory of Scattering of Radio Waves in the Troposphere and Ionosphere

By H. G. BOOKER
Cornell University

Abstract

Some success has recently been achieved by Booker and Gordon [1] in applying a theory of atmospheric scattering to explain certain phenomena of tropospheric propagation, and the same theory, with appropriate modifications, has now been applied to the ionosphere. It throws light on the phenomena of scattering in the E region, and could no doubt be used in connection with auroral phenomena and with scattering in the F region at times of ionospheric storms. The theory confirms an earlier suggestion by Booker and Wells [2] that the cut-off frequency for E_s is controlled more by the size of the fine structure in the E region than it is by the maximum electron-density. The scale l of the fine structure (as used in the theory of turbulence) is about $1/(4\pi)$ times the wave length at which the E_s echo disappears. The strength of this echo should, according to the theory, be independent of wave length down to about the wave length $4\pi\, l$, and its electric field should then decrease proportional to the square of wave length. The attenuation of the transmitted wave due to scattering should likewise be independent of wave length down to about the wave length $4\pi\, l$ and should then be proportional to the square of the wave length. To fit the theory to observations, statistical departures of the electron density from mean are required that vary from a few per cent up to some thirty per cent. The maximum usable frequency for communication to a distance by means of E_s is, fortuitously, very similar to what would be expected if the cut-off frequency of E_s were interpreted in terms of maximum electron density instead of in terms of the scale of the fine structure.

It is an important consequence of the theory that disappearance of the E_s echo as the wave length decreases through the value $4\pi\, l$ approximately is due to a change from roughly omnidirectional scattering to predominately forward scattering, and that forward scattering is practically independent of frequency. This means that there is as much forward scattering at wave lengths less than 10 meters in the E region as there is backward scattering below the cut-off frequency of E_s. Scattering in the E region (and probably also the F region) should therefore make an important contribution to fading of radiation entering the earth's atmosphere from cosmic sources. This fading will be most pronounced on cosmic sources which subtend an angle of less than $\lambda/(2\pi\, l)$, which is the angle, measured from the direction of incidence, within which the forward scattering is

mainly confined. For such a source the ratio of the fading range to the mean field should be proportional to the wave length, and to the square root of the secant of the zenith angle of the source. Observations of the cosmic source in Cygnus by Bolton and Stanley [3] and by Seeger seem to fit in with the idea that an important part of the variation of this source is due to "ionospheric twinkling."

Bibliography

1. Booker, H. G., and Gordon, W. E., *A theory of radio scattering in the troposphere*, Proceedings of the Institute of Radio Engineers, Volume 38, 1950, p. 401.
2. Booker, H. G., and Wells, H. W., *Scattering of radio waves by the F-region of the ionosphere*, Terrestrial Magnetism and Atmospheric Electricity, Volume 43, 1938, p. 249.
3. Bolton, J. G., and Stanley, G. J., *Variable source of radio frequency radiation in the constellation of Cygnus*, Nature, Volume 161, 1948, p. 312.

Properties of Guided Waves on Inhomogeneous Cylindrical Structures

By R. B. ADLER
Massachusetts Institute of Technology

Abstract

An analysis is given of some basic properties of exponential modes on passive cylindrical structures, in which ϵ, μ and σ vary over the cross section and the bounding surface is not completely opaque. Major, but not exclusive, consideration is directed to lossless structures. Each mode is generally a *TE-TM* mixture. Some of the conventional orthogonality conditions do not remain valid. Conditions are discussed under which the instantaneous-, vector-, or double-frequency power flows along the structure are additive among the modes. Stored and dissipated energies generally are not additive. It is shown that the propagation constant for modes on a lossless structure cannot be complex; when the lossless structure has no confining boundary (like a dielectric rod), the modes cannot even possess a true cutoff. Consideration is given to the relation between the direction of real power flow and that of the phase and group velocities. The frequency dependence of the field distribution is also interpreted.

Further information on this material can be found in Technical Report No. 102, May 27, 1949, of the Research Laboratory of Electronics, M.I.T.

Evaluation of Integrals Associated with Wave Motion in Dispersive Media and the Formation of Transients

By MANUEL CERRILLO
Massachusetts Institute of Technology

Abstract

(*a*) One basic research project of the Research Laboratory of Electronics, M. I. T. is the investigation and development of methods of approximate integration of a class of integrals of the type

$$f(\tau_1, \cdots, \tau_n) = \frac{1}{2\pi i} \int_\gamma F(s) \exp \{W(s, \tau_1, \cdots, \tau_n)\}\, ds, \tag{1}$$

particularly the subclass associated with transient phenomena in electrical networks and other linear systems.

Several methods have been investigated and developed. They can be classified into three basic groups: Open methods, Cliff methods and Pocket methods. This classification is in accordance with the way in which the contour of integration, γ, encircles the singularities or passes in their vicinity.

Other subclassification has been made depending on the method in which we approximate the function $F(s)$, or $W(s, \tau_1, \cdots, \tau_n)$ or both. These subclassifications are: stationary phase, saddle point, extended saddle point, plain and mixed cliff, plain pocket, essential and substitutional methods. They are listed in the order of increasing generality. The last method is by far the most complete theoretically. It has been found that the substitutional method possesses powerful potentialities because it is a sort of generalization of the preceding ones.

Investigation of this last method began about a year ago in a rather limited extent. Even at that time extensive applications of the method were foreseen in connection with the synthesis problem, etc., and as a consequence considerable attention has been given to the method during the last few months.

(*b*) The basic ideas contained in this powerful method are outlined here in a condensed form.

We are all aware that the singularities of $F(s)$ and $W(s, \tau_1, \cdots, \tau_n)$, their branch cuts, etc., have a canonical role in the genesis of the integral formation. With certain conditions on the contour γ, which type (1) may satisfy, it happens that the complete integral contribution comes directly from the neighborhood of the singularities of the integrand and along the banks of the cuts, around which we deform the contour of integration.

The theoretical experience with these integrals indicates that the net quantitative effect of the singularities on the integral solution depends strongly on their nature and more strongly still upon their relative position in the s-plane. Because of the relationship of the parameters, $\tau_1, \cdots, \tau_n$ and the functions in the integrand, it becomes clear that these parameters influence, in general, the position, or sometimes the nature, of the above-mentioned singularities.

In our integrals, the parameters $\tau_1, \cdots, \tau_n$ possess a variable nature (independent, of course, of s). When they change, the singularities of the integrand may move in the s-plane, changing their relative position, or sometimes their nature, and therefore changing their effect on the process of the integral formation.

We may naturally wonder if this motion of the singularities of the integrand as a function of $\tau_1, \cdots, \tau_n$ can be exploited as a possible basic idea for a method of approximate integration. To illustrate this possibility we must describe the process in more precise terms.

Suppose that we are interested in an approximate solution of an integral contained in class (1), which will be valid in a certain domain of variation of $\tau_1, \cdots, \tau_n$, say G_ν.[1] When $\tau_1, \cdots, \tau_n$ vary in G_ν, the singularities of the integrand will, in general, move in the s-plane. They follow certain orbits with a definite law of motion, thus changing their relative position. It may happen that during relative motion a certain group of singularities have, or may attain, an almost dominant quantitative control on the building up of the integral solution when $\tau_1, \cdots, \tau_n$ are in G_ν; while other singularities have, or may attain, a small or secondary quantitative effect on the integral when $\tau_1, \cdots, \tau_n$ are in G_ν. Let us assume momentarily that this is the case.

The following notation is now convenient. Let $\tau_1, \cdots, \tau_n$ be in G_ν. "Primary" singularities (in G_ν) are those which have a strong quantitative influence and "secondary" singularities (in G_ν) are those which have a minor quantitative effect on the integral solution.

Now, let us follow the motion in the s-plane of the singularities of the integrand, with particular attention given to the primary ones. Let us trace in the s-plane the orbits and find their laws of motion,[2] in terms of $\tau_1, \cdots, \tau_n$, for each singularity.

The segments of the orbits, corresponding to the displacements of the primary singularities for $\tau_1, \cdots, \tau_n$ in G_ν, will occupy one or several regions of the s-plane, D_ν. The union of these s-plane regions will be denoted by D_ν.

---

[1]We can think of G_ν as follows: the n-tuple of variable numbers $(\tau_1, \cdots, \tau_n)$ defines an n-dimensional vector space, say R_n, over the field of definition of the numbers $\tau_1, \tau_2, \cdots, \tau_n$. Hence G_ν is a subdomain of R_n.

[2]The reader may have a clearer picture of this process if we particularize the general case by considering that $\tau_1, \cdots, \tau_n$ are, for example, continuous functions of the real variable t and set

$$\tau_k = \tau_k(t); \qquad k = 1, \cdots, n$$

in the interval $t_a < t < t_b (G_\nu \text{ is } t_b - t_a)$. The above equations define a line immersed in R_n.

This situation clearly establishes a certain correspondence (not necessarily one-to-one) between G_ν and one, or several regions of the s-plane, D_ν .

The final wave shape produced by the contributions of the primary singularities strongly depends upon the combined effect of every primary singularity. We may therefore look at them as a group rather than as isolated singularities. This group consideration is important particularly because: (a) For a given set and disposition of primary singularities, the relative displacement of one of them may cause, and this is often the case, a considerable change of the wave shape of the integral solution. (b) In some cases we can replace the group by some other analytical entities[3] of simpler structure whose effect with regard to the integral solution is equivalent or almost equivalent, and vice versa. These entities sometimes define points in the s-plane which also move as functions of $\tau_1, \cdots, \tau_n$.

We will agree that when at least one of the primary singularities of a given group drops to a secondary rank, or as soon as at least one new singularity rises to the primary rank, then the given group must be considered as a different one.

(*c*) Past experience with the approximate integration of a large subclass of integrals of type (1) has revealed and confirmed:

(a) The reality of the primary and secondary singularities associated with a given domain G_ν, of variation of $\tau_1, \cdots, \tau_n$.

(b) The raising or lowering in rank of certain singularities in different domains, say G_ν and $G_{\nu+n}$.

(c) The decisive influence on the wave form of the primary singularities when they are considered as a group rather than in their individual roles.

(d) A basic structural composition of the primary group. For example, the improper modification of the group elements, say, by cancellation of one or several singularities, alteration of orbits or laws of motion, may produce an incorrect integral approximation, or a very slowly convergent solution.

(*d*) In the light of all these observations, the following two questions arise. Let $F_\nu^*(s)$ and $W_\nu^*(s, \tau_1, \cdots, \tau_n)$ be two functions which satisfy the requisites:

(a) They approximate respectively $F(s)$ and $W(s, \tau_1, \cdots, \tau_n)$, i.e.

$$\left.\begin{aligned} F_\nu^*(s) &\sim F(s) \\ W_\nu^*(s, \tau_1, \cdots, \tau_n) &\sim W(s, \tau_1, \cdots, \tau_n) \end{aligned}\right\} \begin{aligned} & s \in D_\nu \\ & \tau_k \in G_\nu ; \qquad k = 1, \cdots, n. \end{aligned}$$

(b) The functions $F_\nu^*(s)$ and $W^*(s, \tau_1, \cdots, \tau_n)$ contain the primary singularities (or corresponding analytical entities) respectively of $F(s)$ and $W(s, \tau_1, \cdots, \tau_n)$ for $\tau_k \in G_\nu$, $k = 1, \cdots, n$; $s \in D_\nu$.

---

[3]Among those entities are saddle points, branch cuts, window functions, etc. Inversely, branch cuts can be replaced by certain sequences of alternated poles and zeros.

(c) Suppose that the integral defined by

$$f_\nu^*(\tau_1, \cdots, \tau_n) = \frac{1}{2\pi i}\int_\gamma F_\nu^*(s) \exp\{W_\nu^*(s, \tau_1, \cdots, \tau_n)\}\, ds$$

exists for $\tau_k \varepsilon G_\nu$, $k = 1, \cdots, n$.

Questions:

I. Are the conditions a to c sufficient to assure that $f^* \approx f$ for $\tau_k \varepsilon G_\nu$, $k = 1, \cdots, n$ (but not necessarily for $\tau_k \not\varepsilon G_\nu$) within a certain set of small tolerances? Supposing an affirmative answer for I:

II. If now the orbits of the primary singular points of F and W and, respectively, the orbits of the singular points of F_ν^* and W_ν^* become closer and closer and the respective laws of motion become more and more nearly equal for $\tau_k \varepsilon G_\nu$, $k = 1, \cdots, n$, then, can we say that the set of tolerances mentioned in I becomes smaller and smaller for $\tau_k \varepsilon G_\nu$ but not necessarily for $\tau_k \not\varepsilon G_\nu$, $k = 1, \cdots, n$?

If one considers as "primary" every singularity of the integrand, then a positive answer may be given to the above questions and the solution is indeed valid for the complete domain of variation of $\tau_1, \cdots, \tau_n$. But if one considers only the primary group, disregarding or modifying the secondary ones, then, the answer may be positive or negative and the problem is open to a further investigation. Direct mathematical proofs are very hard to conceive and construct.

In view of some indirect consequences and results, which are derived from previous, known theories of approximate integration, we are fully aware, at least in many specific but illuminating cases, that we may affirm the possibility of a positive answer. The illuminating cases above suggest the existence of a positive answer for many other cases.

The above sequence of ideas serves as a possible base for the development of a method of approximate integration. The above ideas, however, are not directly constructive since: (1) No means are provided to find or test the primary singularities corresponding to a given G_ν. (2) No methods are given to construct the functions F_ν^* and W_ν^*. (3) It is presupposed that the integral

$$\frac{1}{2\pi i}\int F_\nu^*(s) \exp\{W_\nu^*(s, \tau_1, \cdots, \tau_n)\}\, ds$$

exists, but no means are provided to perform the corresponding integration.

An intensive research has been conducted by the author, particularly during the last year, in order to supply the constructive means which are necessary in applying the above ideas. The new method of integration along these lines was called "substitutional." It is fairly well advanced, although it is still far from a final goal. To illustrate this we give below a few examples of the approximate integration of a subclass of integrals. The results are striking for their simplicity.

Let

$$f(t) = \frac{1}{2\pi i} \int_{\gamma} F(s) \exp \{st - \phi(s)\}\, ds; \qquad n = 1,\ \tau_1 = t,$$

where

$$\int_{\gamma} = \int_{c_0 - i\infty}^{c_0 + i\infty},$$

c_0 being the abscissa of uniform convergence.
Let us consider solutions for small values of t in the cases:
Case I:

$$F(s) = C \frac{\prod_{1}^{n} (s - \overset{0}{s}_k)}{\prod_{1}^{m} (s - \overset{x}{s}_i)}; \quad \phi(s) \equiv 0; \quad \text{let } p = m - n; \quad q_0 = \sum_{1}^{n} \overset{0}{s}_k - \sum_{1}^{m} \overset{x}{s}_i$$

Case II:

$$F(s) = C \frac{\sqrt[\mu]{\prod_{1}^{\alpha} (s - \overset{0}{s}_\mu)}\ \prod_{1}^{\delta} (s - \overset{0}{s}_h)}{\sqrt[\nu]{\prod_{1}^{\beta} (s - \overset{x}{s}_\nu)}\ \prod_{1}^{\gamma} (s - \overset{x}{s}_k)}; \qquad \phi(s) \equiv 0;$$

let $p = \gamma - \delta + \dfrac{\beta}{\nu} - \dfrac{\alpha}{\mu}$; $q_0 = \sum_{e} \overset{0}{s}_e j_e - \sum_{m} \overset{x}{s}_m j_m$;

j_k algebraic multiplicity of the pole or zero.

Case III:
$F(s)$ as in cases I and II; $\phi(s) \not\equiv 0$ but satisfying condition $\phi(s) \underset{s \to \infty}{\longrightarrow} Ms^N$ with $N \leq 1$.

Let $\phi(s)$ be expanded around $s = \infty$ as

$$\phi(s) = s\left(a_0 + \frac{a_1}{s} + \frac{a_2}{s^2} + \cdots\right).$$

General Solution (three cases):

$$f(t) \sim u_{-1}(\tau) \exp \{-a_1 \tau\} \frac{\tau^{p-1}}{\Gamma(p)} \Lambda_{(p-1)}(X); \qquad 0 \leq \tau \ll 1$$

where:

$u_{-1}(\tau)$ = unit step function

$\Lambda_{(p-1)}$ = lambda function of order $(p - 1)$

$$\tau = t - a_0; \qquad X = 2\sqrt{\tau(q_0 + a_0)}$$

$$\left.\begin{array}{l}\text{Case I} \\ \text{Case II}\end{array}\right\} \begin{array}{l} a_0 = 0 \\ a_1 = 0 \end{array}$$

Note: one single-term solution for this large family of integrals.

(*e*) Basic ideas of the particular methods of integration, which are mentioned in section (*a*), as well as illustrative examples, were given in the Symposium presentation of this paper. A complete and detailed discussion of the subject will appear in a Technical Report No. 55, M.I.T., Research Laboratory of Electronics.

Electromagnetic Research in the U. S. Air Force Research Program

By N. C. GERSON
Geophysical Research Directorate, Air Force Cambridge Research Laboratories

In a few pages, it is not possible to describe in detail a full program of mathematical investigations in electromagnetics which the U. S. Air Force would like to see pursued. However, some particular fields of interest may be briefly and generally indicated.

At the risk of repetition, I would like first to state the Air Force policy with respect to research. The Air Force, in fostering studies in mathematical physics, does so with the clear conviction that today's research becomes tomorrow's practice, and that any policy which confines itself solely to present problems and their immediate solution soon leads to scientific degeneration and stagnation. It is true that the Air Force desires some applications of its sponsored research. However, the obtainment of maximum results requires that this view be liberally interpreted in the sense of forging tools for mathematical physics which will be available not only now, but also, if required, ten, twenty, or fifty years hence. In applying this policy towards investigations in electromagnetics, the Air Force finds itself vitally concerned with long-term studies, such as extension of present theory and clarification of anomalies.

During the war, the necessity for mathematical investigations in electromagnetics was obvious. The results were phenomenal. As one example, we may mention the advances made by Schwinger's integral equation and variational techniques, and the impact made by these on contemporary research in pure and applied physics. The celebrated success of concentrated investigations in mathematical physics, obtained by such groups as Radiation Laboratory, Massachusetts Institute of Technology, and the Institute for Mathematics and Mechanics of New York University, are well established. The present Air Force sponsorship of electromagnetic research at New York University represents not only a continuation of the wartime initiated studies, but also a marked expansion in their scope and objectives.

In jointly sponsoring this Conference, both the AF Cambridge Research Laboratories and the New York University had but one objective in mind: the dissemination of recent advances in electromagnetics among the participants and the infusion of new vigor into current research. Because of recent progress in some aspects of the science, it becomes opportune to review at this time the advances in electromagnetics, and to examine the hypotheses upon which they stand. Additional progress may still be made in utilizing unexploited techniques for applied problems. It is wise to uncover the stumbling blocks in electromag-

netic research and to search for means to bypass them. From an open discussion of the results and pitfalls, new thoughts arise, and the general stimulation breeds further progress in this field.

There are innumerable problems in electromagnetics still awaiting solution. It is the hope of the sponsors that a partial listing will stimulate interest in and promote thinking about these problems. Indeed, it is very possible for the researcher on the scene to overlook the obvious and become entrapped in more complicated solutions. It is hoped that the various problems outlined below will germinate and sprout in the minds of some of the readers.

The subjects of interest to the Air Force are not all being investigated at the present time; even if they were, reproducibility of results is desirable. The various research topics fall into several distinct but interlacing categories and may be described as: (a) development of broad mathematical techniques and procedures; (b) study of theoretical problems having present application; and (c) research on the foundations and rigorous development of electromagnetic theory.

The first classification covers a thorough investigation of new mathematical devices needed for the solution of the various integral equations of electromagnetics. In order to make significant progress in any phase of activity, proper equipment is necessary; in electromagnetics, equipment is synonymous with mathematical techniques. In this connection, a considerable amount of additional effort is yet required on phase integral methods; these methods may then be employed to obtain approximate solutions of the differential equations not only of wave mechanics, but also of propagation and diffraction. The W.K.B. method and its possible applications to partial differential equations should be more completely investigated. The usual W.K.B. procedure neglects reflections; however, by modified methods, solutions, $\phi(x)$, may be obtained in terms of an infinite series of the type

$$\phi(x) = \phi_0 + \phi_1 + \phi_2 + \cdots$$

The first term, ϕ_0, represents the common W.K.B. approximation; the next term, ϕ_1, represents the distributed reflections ignored by the first term; the third term, ϕ_2, represents the reflections produced by the second term, etc. Since the occurrence of reflections gives rise to forbidden zones, the W.K.B. method does not provide the zone structure required for periodic potentials. By considering the first order reflection term, ϕ_1, in the above series, however, a simple description of band structure may be obtained. Other extensions are needed particularly with respect to low values of electromagnetic frequencies and complex dielectrics.

The widely known basic method applied and adapted by Schwinger employs variational techniques for the solution of a variety of physical problems. With this method the function that makes a given ratio of integrals stationary is the same function as the solution of a certain integral equation. The standard procedure, which has been employed to obtain solutions of the vector wave equations,

lies in (a) formulating the problem as a differential equation with specific boundary conditions; and (b) transforming it through the use of Green's theorem into an integral equation. By utilizing Schwinger's variational theorem and attempting suitable trial functions in the variational problem, a solution may be obtained, and, in some cases, an estimate of the error involved.

Obtaining the estimated error is a great advantage of Schwinger's work when compared to other methods of geometrical optics or of successive approximations. The method has been employed to determine the diffraction of a scalar plane wave by an aperture in an infinite plane screen, and to study proton-neutron scattering at low particle energies. Application of variational techniques has been made to quantum mechanics in an effort to determine accurate values of the asymptotic neutron densities under given conditions. It seems reasonable to expect that Schwinger's variational techniques may be used advantageously to solve more difficult aspects of specific vector wave problems by means of dyadic Green's functions.

Another route for exploration lies in determining those coordinate systems wherein the scalar and the vector wave equations are separable. This general problem has been extensively studied by the pure mathematicians, and some progress has been made on the special problem of separating variables of the one particle Schroedinger equation in three-space, and in the separation of variables for the two particle wave equation. The fruitful results already obtained in seeking special coordinate systems wherein the wave equation is separable encourage further study.

Extension of the Weiner-Hopf technique, required for the exact solutions of some diffraction problems, should also be undertaken. With this method, Green's theorem or a modal analysis is employed to represent the solution by means of an integral equation, which is then solved by a Fourier transformation and the process of analytic continuation.

Additional mathematical techniques which hold promise may also be mentioned. Studies on the expansion of solutions of Maxwell's equations in orthogonal functions may have useful applications in theoretical quantum electrodynamics as well as in theories of microwaves. This technique frees theory from the requirements that the field be expanded in plane waves. Another method to be investigated is that of determining alternative representations of Green's functions which, in general, lead to a more direct and useful solution of the wave equation. It is highly desirable on some occasions to represent the solution of a differential equation in terms of different series, each having a different domain of rapid convergence.

With respect to the second category of interest to the Air Force, i.e., the host of problems which have some practical applications, only a few can be mentioned; these few are in the field of propagation and scattering of electromagnetic waves. A very fertile field for increased investigative activity is presented by theories of diffraction. Diffraction by almost any object, including circular discs still requires considerable study, particularly with respect to edge effects in the

immediate vicinity of the diffracting object. Rayleigh pointed out that certain singularities occur at the sharp edges of the diffracting screen, a principle later demonstrated in Sommerfeld's solution of diffraction by a semi-plane. Diffraction of electromagnetic waves around a sphere, considered rigorously by Mie and extended by Wilson, has recently been reconsidered by Foch. For his work Foch assumed that the transition from the illuminated to the shadow region on the surface of a sphere occurred in a narrow strip along the boundary of the geometrical shadow. These investigations, however, are but a beginning, and more work should be attempted. One problem of great complexity is that of diffraction by a dielectric wedge. The general problem is as yet unsolved although various approximations have been proposed.

A little progress has already been made in treating diffraction by a random screen, but the topic as a whole, including diffraction by a sphere whose surface is periodically perturbed, has scarcely been examined. These problems are generalities of specific topics which arise in connection with electromagnetic propagation around obstacles, with radio wave propagation through meteorological frontal systems, and with scattering of radio waves by ionospheric or tropospheric inhomogeneities. Simple aspects of some of these topics have already been partially treated.

The scattering of electromagnetic waves by dielectric discontinuities requires further amplification, especially for such problems as scattering by ellipsoids. Investigations must also be undertaken on the scattering of waves by plane surfaces (such as hexagonal plates), considering variously polarized waves as well as specific orientations of the plane of the scatterers. The problem should not only be considered for plane surfaces of constant size, shape and dielectric constant, but also under conditions wherein each of these factors is randomly distributed about a mean value. Scattering by means of spherical and spheroidal particles whose dielectric constant is a function of both space and time also requires examination. These and similar problems arise in microwave meteorology in connection with the scattering of microwaves by snowflakes and ice particles, both of which may melt as they fall to the ground.

Propagation problems may by no means be neglected. In this group are those involving wave propagation (at appropriate radio frequencies) through the ionosphere, the troposphere, and through very highly ionized media, as, for example, may be found in stellar atmospheres. Propagation over rough surfaces, such as irregular terrain, or over rough time-varying surfaces, as over swaying trees or a disturbed ocean should be given more attention. With respect to the ionosphere, better methods of determining true heights of the reflection layer are desired. Solutions of the wave equation near the "reflection" surface for very long radio waves refracted from a medium (the ionosphere) where the refractive index is a rapidly varying function of the coordinates are still lacking. Additional thought is also necessary on non-linear propagation in the ionosphere, as well as on wave propagation through a highly absorbing volume (exemplified by the auroral zone).

The third category for study may be termed special topics in electromagnetic theory. The use of asymptotic expansions described at this symposium in the paper, "Asymptotic Expansion of Electromagnetic Fields," by Dr. M. Kline, warrants considerably more extension. Increased studies are needed to determine appropriate singularities in the electromagnetic field near sharp, well defined corners and edges.

It is perhaps appropriate to mention that in addition to Maxwell's classic equations, Gibbs' fundamental work on dyadic analysis has led to an adequate account of wave reflection and absorption in both isotropic and anisotropic media, and to correct results for dispersion and absorption. Although Gibbs' work is in his usual comprehensive style, his advances in this field have received somewhat scant attention; theoretical work utilizing his theorems still merits serious study.

The reader may perhaps wonder at the Air Force's interest in such a large variety of electromagnetic studies. The first group of investigations is devoted to the tooling stage, i.e., to the derivation of techniques and devices; the second category considers applications; and the third encompasses special problems. Our interest in all groups arises from the fundamental premise, proven by the historical development of the physical sciences, that the *research of today is the practice of tomorrow*, and that unless we devote an absolute minimum of activity to long-range fundamental studies, we have no investment in the technology of the future.

Research in the United States today represents an unbalanced program. Great effort is devoted toward experimental investigations whereas we lag dangerously in theoretical studies. Must we wait for the pressure of an emergency to develop our theories and techniques, or is it possible during normal periods to theorize and progress? Why can we not balance our effort, not by dropping our experimentation, but by increasing our theoretical efforts? Certainly there is much more to research than instrumentation or the immediate solution of particular applied problems.